

**Pages 1 to / à 2**  
**are withheld pursuant to section**  
**sont retenues en vertu de l'article**

**23**

**of the Access to Information Act**  
**de la Loi sur l'accès à l'information**

The Agile Auditor in the Age of Disruption  
L'auditeur agile dans une ère de perturbation

# ***Creating an Enterprise-Wide View of Risks Through Deep Learning***



NATIONAL CONFERENCE  
CONFÉRENCE NATIONALE

Le Centre Sheraton Montréal Hotel | September 30 - October 3, 2018



# Employment and Social Development Canada Context



3 Million applications processed for Employment Insurance

27 Million payments issued for Employment Insurance



4.8 Million passports issued



689,764 applications processed for Canada Pension Plan and 800,941 for Old Age Security

66.3 Million payments issued for Canada Pension Plan and 70.5 Million for Old Age Security



\$3.56 Billion was withdrawn from Registered Education Savings Plans, supporting post-secondary education

562,000 full-time post-secondary students (aged 15-29) received federal student financial assistance

The Agile Auditor in the Age of Disruption / L'auditeur agile dans une ère de perturbation



NATIONAL CONFERENCE  
CONFÉRENCE NATIONALE

Le Centre Sheraton Montreal Hotel | September 30 - October 3, 2018

# ESDC Data Strategy



VISION

*"Everyone has access to the data they need when they need it"*

MISSION

Provide people easy, secure, and authorized access to quality data in a way that respects personal privacy and delivers value by giving them the skills, tools, and processes needed to maximize the impact of our enterprise data asset.

DRIVERS

Results & Delivery

Policy Mandate, Research & Evaluation

Client Service & Transformation

Internal Services

PRINCIPLES

Client First

Know & serve our clients better

Trust

Privacy, Ethics & Data for the right purpose

Agility

Technology & Innovation

Collaboration

Align to generate exponential value

GOALS



1 People have knowledge & tools to use data



2 Data is a corporate asset through effective governance & stewardship



3 Data is accessible & secure



4 Integrated and streamlined partnership between business & IT



5 Data science & analytics are core skills across ESDC

WHAT ARE WE DOING?

CRAWL

WALK

RUN

	Employees have an understanding, awareness and skills needed to drive the data strategy	Employees as data ambassadors Adoption strategies with key stakeholders	Robust data capacity across ESDC Self-organizing communities solve business problems with data e.g., virtual crowd sourcing
ENABLEMENT	ADKAR adoption plan   First virtual Community of Practice   Chief Data Office Data Literacy Plan	ESDC ADKAR and Data Literacy Plan   Data Skills Incubator   Awareness	ESDC Data Literacy Curriculum   Multi-channel communication   Library/Curation
PEOPLE	Clear path to talent & advance key internal partnerships Chief Data Office HR Plan   Define engagement model   Build internal partnerships	We have the right people to reach our goals Expand internal partnerships & unify data strategies across Government of Canada Chief Data Office function established   Development of ESDC HR Plan and recruitment strategy   Data strategy alignment with key partners   External Stakeholder Network	ESDC's people & partnerships maximize the value of its data Departmental HR Plan and recruitment strategy established   Government of Canada collaboration on recruitment and retention of data talent   Collaboration on key data initiatives (e.g., Children's Data Strategy)
DATA GOVERNANCE	Data governance pockets in place Key data stewards identified Roles & responsibilities   Chief Data Office Data Governance   Data Frame	Overarching governance policies across ESDC Data stewards network supporting implementation Data Quality proof of concept   Enterprise Data Model	Enterprise-wide data governance & stewardship in critical domains Data issues managed in business Data Frame & Principles applied to all data activities   Policy & Standards review & modernization   Aligned data, info and analytics governance
ACCESS	Data access challenges are well understood Data classification understanding   SSPB Open Data lead   Roadmap for the Transfer of ESDC Administrative Data Files to Statistics Canada   Data Access Working Group   Partner to enable Hackathons	Data is accessible for most common scenarios Enhanced data access processes   Partner with Academic Researchers   Research Data Centres support exploratory data work   Public demand driven open data releases	ESDC data is accessible by default Enhanced Data Sharing legislative and policy framework   Enhanced Memorandum of Understanding, Information Sharing Agreements   Comprehensive admin data in Research Data Centres network   Government of Canada & public Open Data search tool
DATA MANAGEMENT	Chief Data Office & Innovation, Information and Technology Branch collaborate on pilots and initiatives while establishing engagement model Establish sandbox   Analytics Reference Architecture   Enterprise Data Architecture Collaboration Model   Data Catalogue Pilot	Durable Chief Data Office - Innovation, Information & Technology Branch collaboration mechanisms Work on transformative data initiatives Solutions Roadmap for Analytics   Technology Innovation Lab Model   Joint Strategy on Self-Service BI   MDM Pilot	Chief Data Office & Innovation, Information and Technology Branch expand scope of collaboration Data Management and Data Governance integration and alignment   Robust ability to provision data and audit access
DATA SCIENCE	Ad hoc data mining & machine learning demo projects Develop hub capacity and services   First analytics pilots   Expanded use of existing tools & technologies   Initiate Department Analytics Program	Test & deploy analytical solutions in operations Maturity-based hub and spoke model   Analytics repository   Strategic in-depth pilots   Data and software exploration   Operationalize Department Analytics Program	Seamless system interaction with analytics & machine learning models Support for priority Enterprise projects   Developed Analytics Agenda   Fully automated integrated Solutions   Complete Analytics Architecture

BUSINESS VALUE & MEASUREMENT — CDO WILL DEVELOP A PLAN FOR MEASURING SUCCESS & CHAMPION THE DELIVERY OF THE DATA STRATEGY

Employment and Social Development Canada (ESDC)

The Agile Auditor in the Age of Disruption / L'auditeur agile dans une ère de perturbation

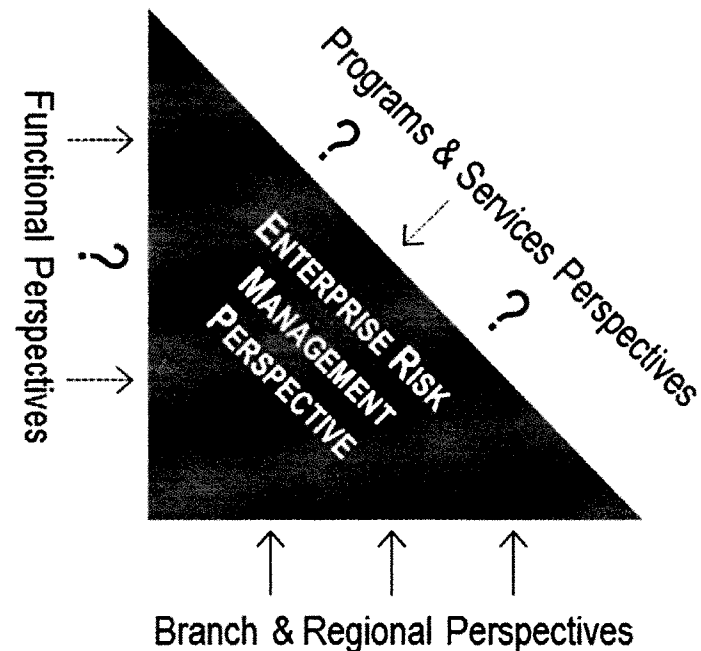


NATIONAL CONFERENCE  
CONFÉRENCE NATIONALE

Le Centre Sheraton Montreal Hotel | September 30 - October 3, 2018

# Problem

- ESDC manages a high volume of risk information spanning branches/regions, functional groups and programs/services.
- It is time-consuming and resource intensive to analyze risk data.
- Centralized perspectives of risk intelligence is fundamental to risk management activities.



The Agile Auditor in the Age of Disruption / L'auditeur agile dans une ère de perturbation

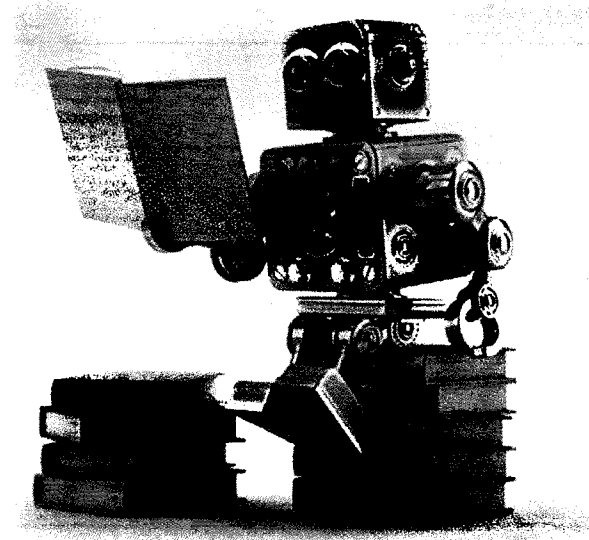


NATIONAL CONFERENCE  
CONFÉRENCE NATIONALE

Le Centre Sheraton Montreal Hotel | September 30 - October 3, 2018

# Potential Solution: Deep learning

- Deep learning enables the recognition of patterns in vast volumes of data that would be impossible for humans to process.
- Deep learning can aggregate and analyze risk information from across the Department to develop new views and insights into risk.
- Strong centralized risk intelligence will provide decision support to enable ESDC to identify and respond to risks.



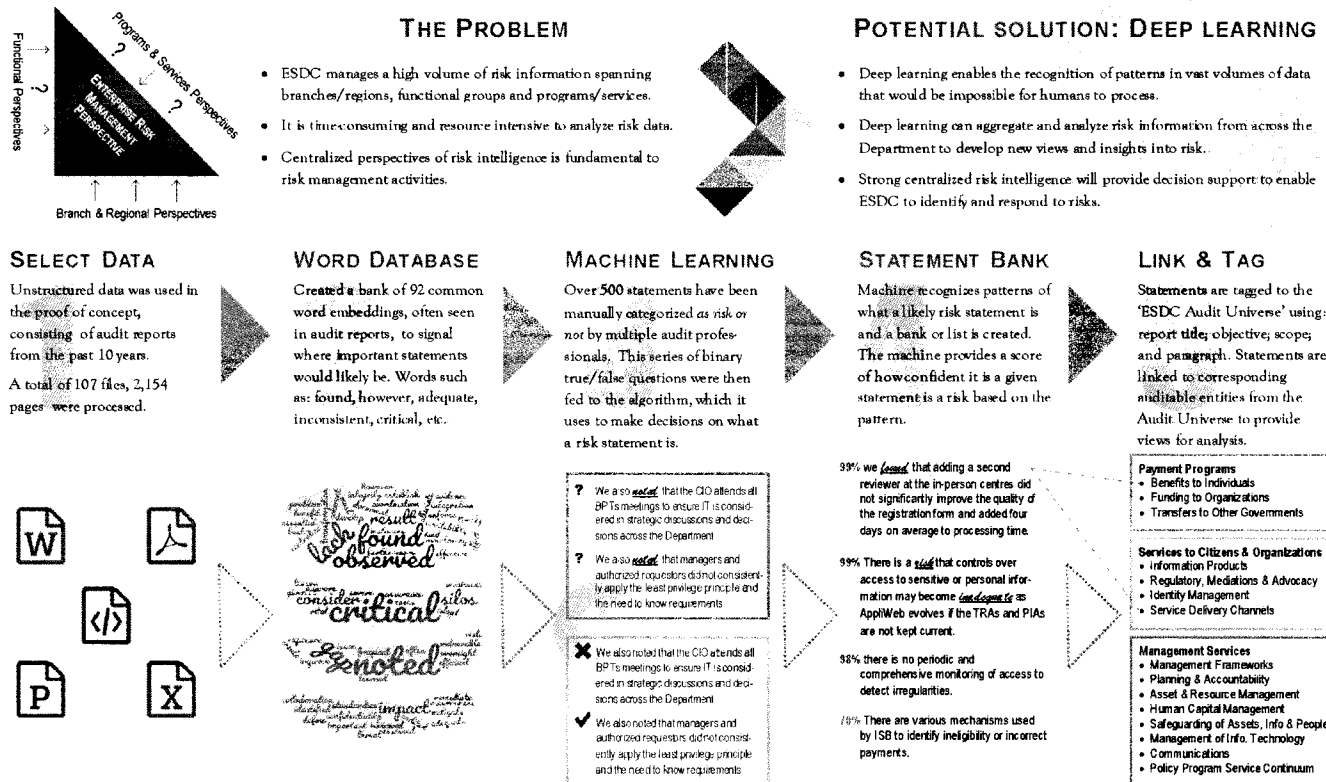
The Agile Auditor in the Age of Disruption / L'auditeur agile dans une ère de perturbation



NATIONAL CONFERENCE  
CONFÉRENCE NATIONALE

Le Centre Sheraton Montreal Hotel | September 30 - October 3, 2018

# Five Phase Project Process



The Agile Auditor in the Age of Disruption / L'auditeur agile dans une ère de perturbation

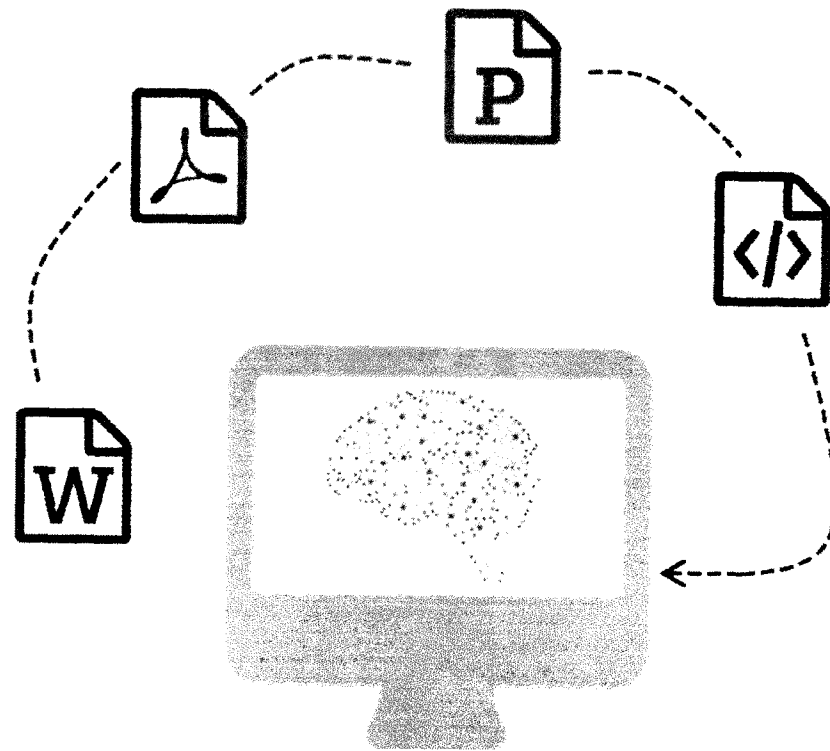


NATIONAL CONFERENCE  
CONFÉRENCE NATIONALE

Le Centre Sheraton Montreal Hotel | September 30 - October 3, 2018

# Select Data

- Unstructured data was used in the proof of concept, consisting of audit reports from the past 10 years.
- A total of 107 files, 2,154 pages were processed.

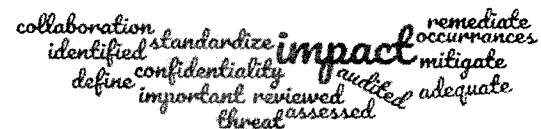
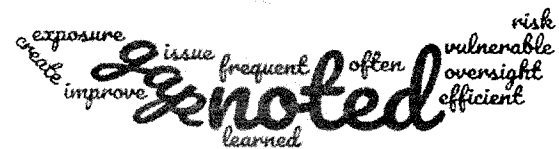


NATIONAL CONFERENCE  
CONFÉRENCE NATIONALE

Le Centre Sheraton Montreal Hotel | September 30 - October 3, 2018

# Word Database

- Created a bank of 92 common word embeddings, often seen in audit reports, to signal where important statements would likely be.
- Words such as: found, however, adequate, inconsistent, critical, etc.



The Agile Auditor in the Age of Disruption / L'auditeur agile dans une ère de perturbation



NATIONAL CONFERENCE  
CONFÉRENCE NATIONALE

Le Centre Sheraton Montreal Hotel | September 30 - October 3, 2018

# Machine Learning

Over 500 statements have been manually categorized as risk or not by multiple audit professionals. This series of binary true/false questions were then fed to the algorithm, which it uses to make decisions on what a risk statement is.

? We also noted that the CIO attends all BPTs meetings to ensure IT is considered in strategic discussions and decisions across the Department

? We also noted that managers and authorized requestors did not consistently apply the least privilege principle and the need to know requirements

✗ We also noted that the CIO attends all BPTs meetings to ensure IT is considered in strategic discussions and decisions across the Department

✓ We also noted that managers and authorized requestors did not consistently apply the least privilege principle and the need to know requirements

The Agile Auditor in the Age of Disruption / L'auditeur agile dans une ère de perturbation



NATIONAL CONFERENCE  
CONFÉRENCE NATIONALE

Le Centre Sheraton Montreal Hotel

September 30 - October 3, 2018



# Statement Bank

Machine recognizes patterns of what a likely risk statement is and a bank or list is created. The machine provides a score of how confident it is a given statement is a risk based on the pattern.

99% we found that adding a second reviewer at the in-person centres did not significantly improve the quality of the registration form and added four days on average to processing time.

99% There is a risk that controls over access to sensitive or personal information may become inadequate as AppliWeb evolves if the TRAs and PIAs are not kept current.

98% there is no periodic and comprehensive monitoring of access to detect irregularities.

70% There are various mechanisms used by ISB to identify ineligibility or incorrect payments.

The Agile Auditor in the Age of Disruption / L'auditeur agile dans une ère de perturbation



NATIONAL CONFERENCE  
CONFÉRENCE NATIONALE

Le Centre Sheraton Montreal Hotel | September 30 - October 3, 2018

# Link & Tag

Statements are tagged to the 'ESDC Audit Universe' using: report title; objective; scope; and paragraph. Statements are linked to corresponding auditable entities from the Audit Universe to provide views for analysis.

99% we *found* that adding a second reviewer at the in-person centres did not significantly improve the quality of the registration form and added four days on average to processing time.

99% There is a *risk* that controls over access to sensitive or personal information may become *inadequate* as AppliWeb evolves if the TRAs and PIAs are not kept current.

98% there is no periodic and comprehensive monitoring of access to detect irregularities.

70% There are various mechanisms used by ISB to identify ineligibility or incorrect payments.

## Payment Programs

- Benefits to Individuals
- Funding to Organizations
- Transfers to Other Governments

## Services to Citizens & Organizations

- Information Products
- Regulatory, Mediations & Advocacy
- Identity Management
- Service Delivery Channels

## Management Services

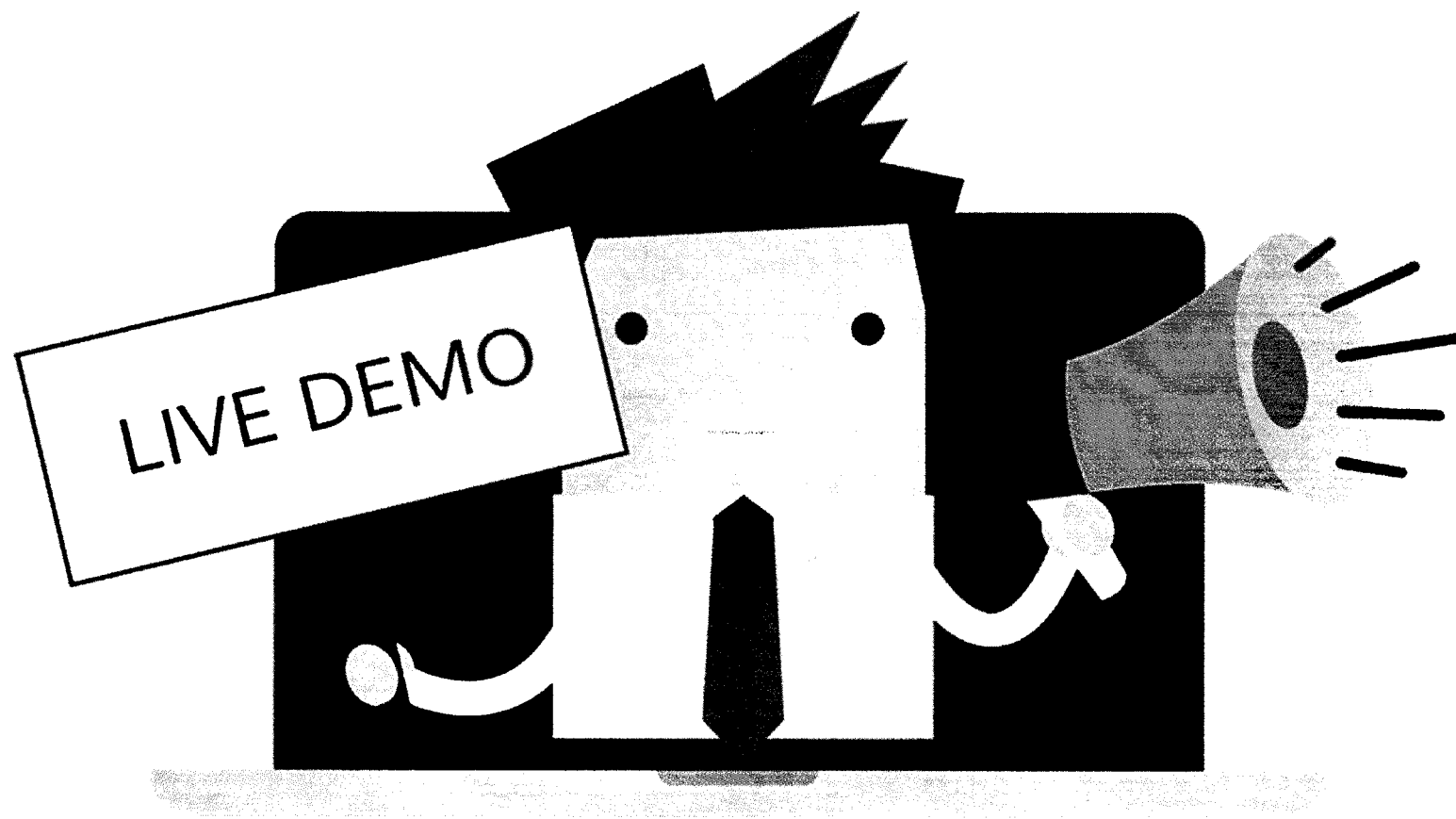
- Management Frameworks
- Planning & Accountability
- Asset & Resource Management
- Human Capital Management
- Safeguarding of Assets, Info & People
- Management of Info. Technology
- Communications
- Policy Program Service Continuum

The Agile Auditor in the Age of Disruption / L'auditeur agile dans une ère de perturbation



NATIONAL CONFERENCE  
CONFÉRENCE NATIONALE

Le Centre Sheraton Montreal Hotel | September 30 - October 3, 2018



The Agile Auditor in the Age of Disruption / L'auditeur agile dans une ère de perturbation

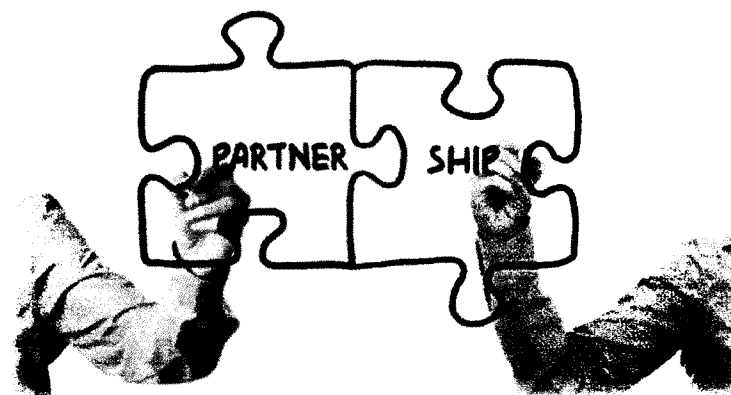


NATIONAL CONFERENCE  
CONFÉRENCE NATIONALE

Le Centre Sheraton Montreal Hotel | September 30 - October 3, 2018

# Lessons Learned - Partnership Matters

- Joint team of business and technology experts
- Data team working with Internal Audit Team to solve problems
- Agile workflow - Enabling feedback in the development process
- Buy vs. Build
- The need for internal technical experts



The Agile Auditor in the Age of Disruption / L'auditeur agile dans une ère de perturbation

# Current Status & Next Steps

- Internal Audit and the CDO continue to advance this innovation beyond the life of the BCIP funding.
- Internal Audit's objective with respect to this machine learning innovation is to determine if the innovation can assemble and analyze risk information across many areas of the Department that could not be easily done manually.
- At present, there is little to no human or IT capacity to analyze risk information in a vertical and horizontal fashion.
- ESDC's Internal Audit and the CDO continue to work on further testing and refining the machine learning / AI solution in-house, which has wider applications across the GOC.

The Agile Auditor in the Age of Disruption / L'auditeur agile dans une ère de perturbation



NATIONAL CONFERENCE  
CONFÉRENCE NATIONALE

Le Centre Sheraton Montreal Hotel | September 30 - October 3, 2018



Employment and  
Social Development Canada

Emploi et  
Développement social Canada

Canada



# **Creating an Enterprise-Wide View of Risks Through Deep Learning**

# Employment and Social Development Canada Context



3 Million applications processed for Employment Insurance  
27 Million payments issued for Employment Insurance



4.8 Million passports issued



689,764 applications processed for Canada Pension Plan and 800,941 for Old Age Security  
66.3 Million payments issued for Canada Pension Plan and 70.5 Million for Old Age Security



\$3.56 Billion was withdrawn from Registered Education Savings Plans, supporting post-secondary education  
562,000 full-time post-secondary students (aged 15-29) received federal student financial assistance



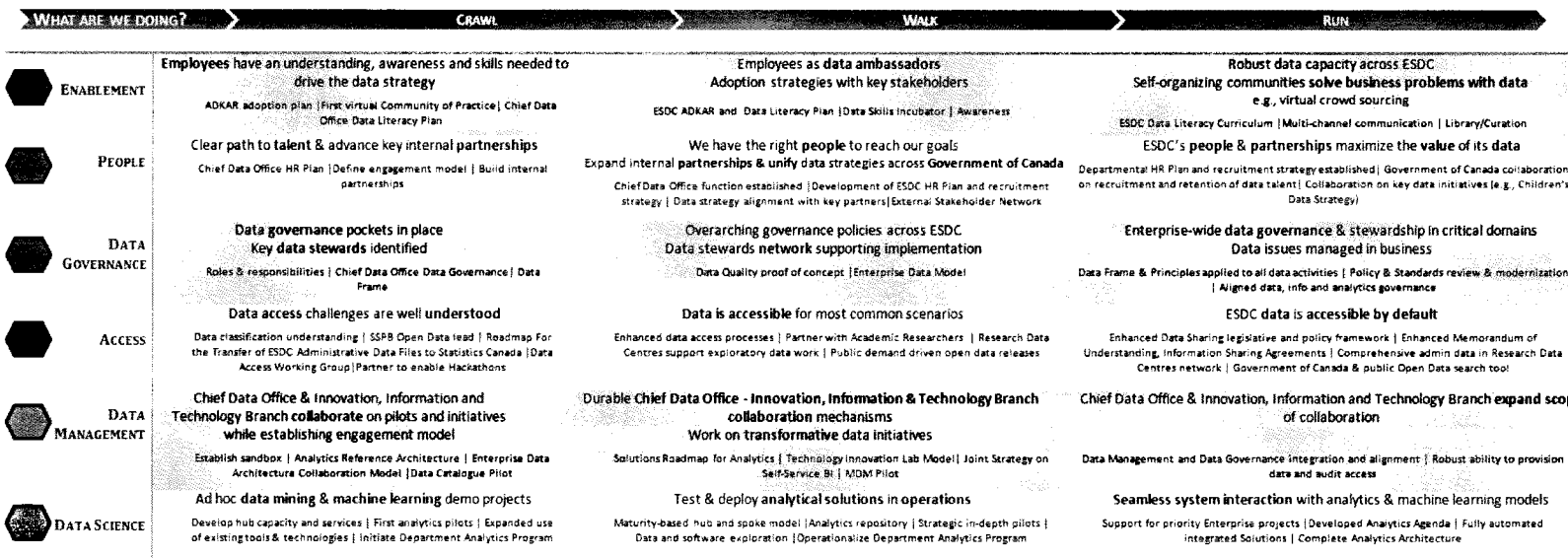
# ESDC Data Strategy



VISION

*"Everyone has access to the data they need when they need it"*

**MISSION** Provide people easy, secure, and authorized access to quality data in a way that respects personal privacy and delivers value by giving them the skills, tools, and processes needed to maximize the impact of our enterprise data asset.



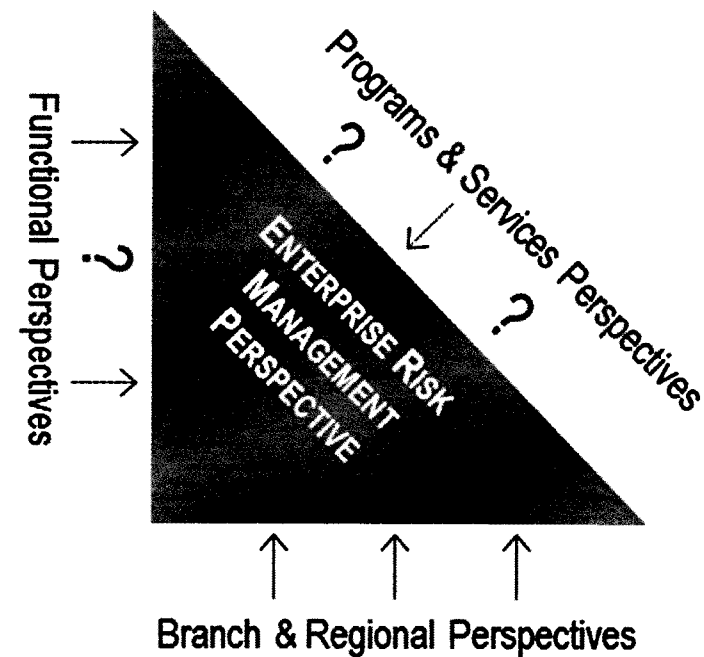
**BUSINESS VALUE & MEASUREMENT — CDO WILL DEVELOP A PLAN FOR MEASURING SUCCESS & CHAMPION THE DELIVERY OF THE DATA STRATEGY**

Employment and Social Development Canada (ESDC)



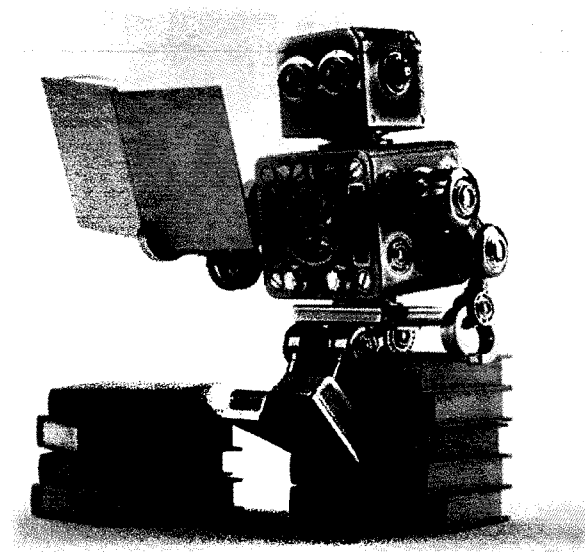
# Problem

- ESDC manages a high volume of risk information spanning branches/regions, functional groups and programs/services.
- It is time-consuming and resource intensive to analyze risk data.
- Centralized perspectives of risk intelligence is fundamental to risk management activities.

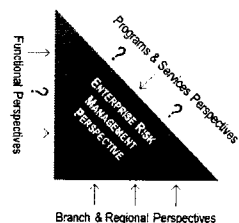


# Potential Solution: Deep learning

- Deep learning enables the recognition of patterns in vast volumes of data that would be impossible for humans to process.
- Deep learning can aggregate and analyze risk information from across the Department to develop new views and insights into risk.
- Strong centralized risk intelligence will provide decision support to enable ESDC to identify and respond to risks.



# Five Phase Project Process



## THE PROBLEM

- ESDC manages a high volume of risk information spanning branches/regions, functional groups and programs/services.
- It is time consuming and resource intensive to analyze risk data.
- Centralized perspectives of risk intelligence is fundamental to risk management activities.

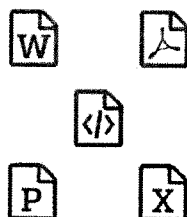


## POTENTIAL SOLUTION: DEEP LEARNING

- Deep learning enables the recognition of patterns in vast volumes of data that would be impossible for humans to process.
- Deep learning can aggregate and analyze risk information from across the Department to develop new views and insights into risk.
- Strong centralized risk intelligence will provide decision support to enable ESDC to identify and respond to risks.

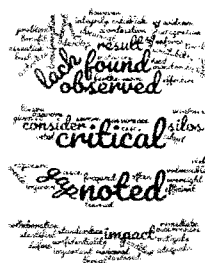
## SELECT DATA

Unstructured data was used in the proof of concept, consisting of audit reports from the past 10 years. A total of 107 files, 2,154 pages were processed.



## WORD DATABASE

Created a bank of 92 common word embeddings, often seen in audit reports, to signal where important statements would likely be. Words such as: found, however, adequate, inconsistent, critical, etc.



## MACHINE LEARNING

Over 500 statements have been manually categorized as risk or not by multiple audit professionals. This series of binary true/false questions were then fed to the algorithm, which it uses to make decisions on what a risk statement is.

- ? We also noted that the QIO attends all BPTs meetings to ensure IT is considered in strategic discussions and decisions across the Department.
- ? We also noted that managers and authorized requestors did not consistently apply the least privilege principle and the need to know requirements.
- X We also noted that the QIO attends all BPTs meetings to ensure IT is considered in strategic discussions and decisions across the Department.
- ✓ We also noted that managers and authorized requestors did not consistently apply the least privilege principle and the need to know requirements.

## STATEMENT BANK

Machine recognizes patterns of what a likely risk statement is and a bank or list is created. The machine provides a score of how confident it is a given statement is a risk based on the pattern.

- 99% we found that adding a second reviewer at the in-person centres did not significantly improve the quality of the registration form and added four days on average to processing time.
- 99% There is a rule that controls over access to sensitive or personal information may become outdated as AppliWeb evolves if the TRAs and PIAs are not kept current.
- 98% there is no periodic and comprehensive monitoring of access to detect irregularities.
- 79% There are various mechanisms used by ISB to identify ineligibility or incorrect payments.

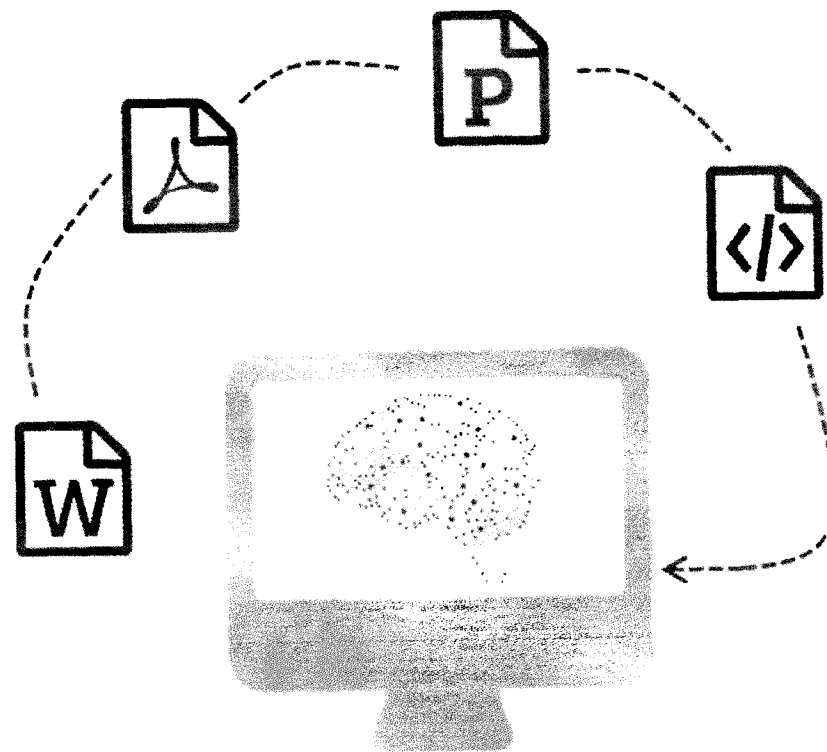
## LINK & TAG

Statements are tagged to the 'ESDC Audit Universe' using: report title, objective, scope, and paragraph. Statements are linked to corresponding auditable entities from the Audit Universe to provide views for analysis.

- Payment Programs**
  - Benefits to Individuals
  - Funding to Organizations
  - Transfers to Other Governments
- Services to Citizens & Organizations**
  - Information Products
  - Regulatory, Mediations & Advocacy
  - Identity Management
  - Service Delivery Channels
- Management Services**
  - Management Frameworks
  - Planning & Accountability
  - Asset & Resource Management
  - Human Capital Management
  - Safeguarding of Assets, Info & People
  - Management of Info. Technology
  - Communications
  - Policy Program Service Continuum

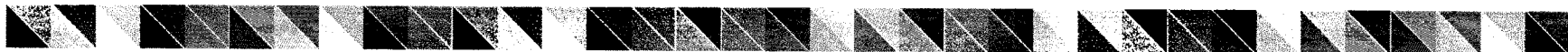
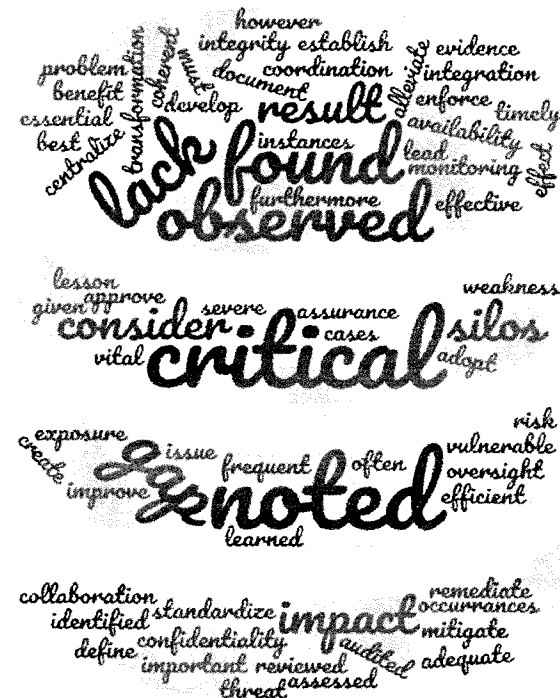
# Select Data

- Unstructured data was used in the proof of concept, consisting of audit reports from the past 10 years.
- A total of 107 files, 2,154 pages were processed.



# Word Database

- Created a bank of 92 common word embeddings, often seen in audit reports, to signal where important statements would likely be.
- Words such as: found, however, adequate, inconsistent, critical, etc.





# Machine Learning

Over 500 statements have been manually categorized as risk or not by multiple audit professionals. This series of binary true/false questions were then fed to the algorithm, which it uses to make decisions on what a risk statement is.

? We also noted that the CIO attends all BPTs meetings to ensure IT is considered in strategic discussions and decisions across the Department

? We also noted that managers and authorized requestors did not consistently apply the least privilege principle and the need to know requirements

✗ We also noted that the CIO attends all BPTs meetings to ensure IT is considered in strategic discussions and decisions across the Department

✓ We also noted that managers and authorized requestors did not consistently apply the least privilege principle and the need to know requirements



# Statement Bank

Machine recognizes patterns of what a likely risk statement is and a bank or list is created. The machine provides a score of how confident it is a given statement is a risk based on the pattern.

99% we found that adding a second reviewer at the in-person centres did not significantly improve the quality of the registration form and added four days on average to processing time.

99% There is a risk that controls over access to sensitive or personal information may become inadequate as AppliWeb evolves if the TRAs and PIAs are not kept current.

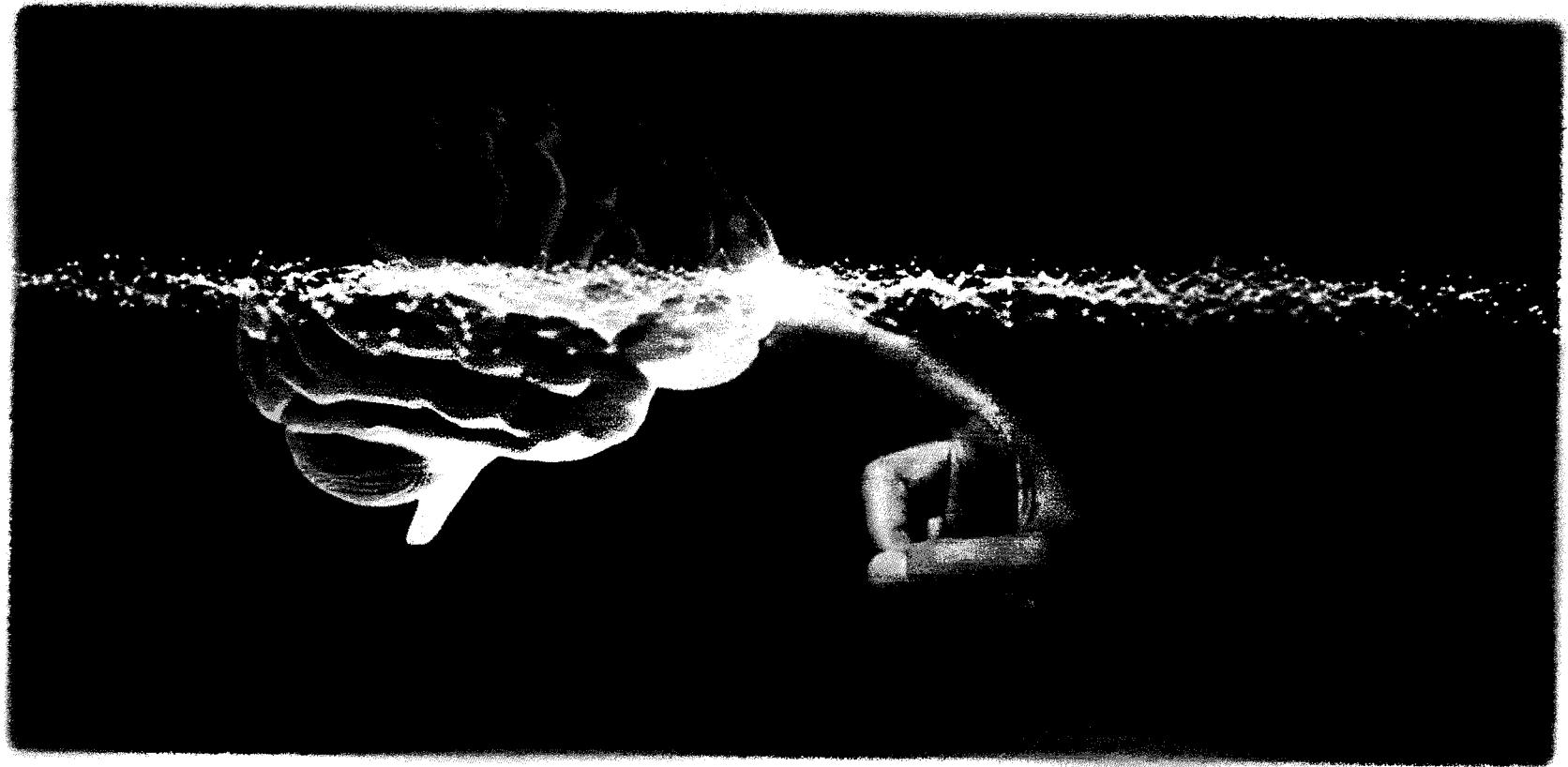
98% there is no periodic and comprehensive monitoring of access to detect irregularities.

70% There are various mechanisms used by ISB to identify ineligibility or incorrect payments.





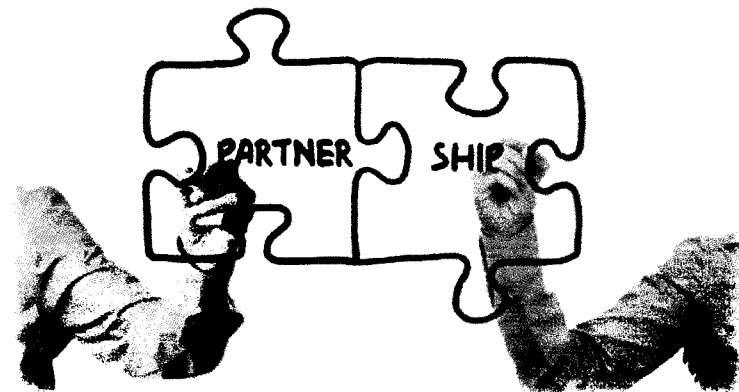
# 3 Visualization Modules





# Lessons Learned - Partnership Matters

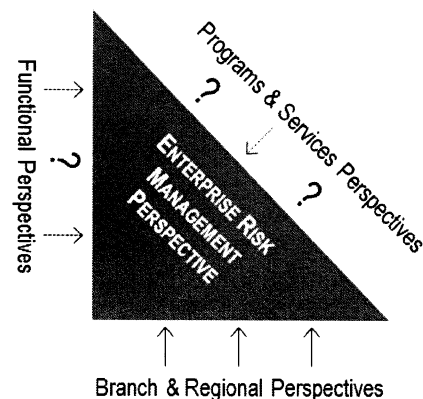
- Joint team of business and technology experts
- Data team working with Internal Audit Team to solve problems
- Agile workflow - Enabling feedback in the development process
- Buy vs. Build
- The need for internal technical experts



# Current Status & Next Steps

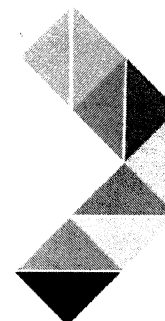
- Internal Audit's objective with respect to this machine learning innovation is to determine if the innovation can assemble and analyze risk information across many areas of the Department that could not be easily done manually.
- Internal Audit and the CDO continue to advance this innovation.
- At present, there is little to no human or IT capacity to analyze risk information in a vertical and horizontal fashion.
- ESDC's Internal Audit and the CDO continue to work on further testing and refining the machine learning / AI solution in-house, which has wider applications across the GOC.





## THE PROBLEM

- ESDC manages a high volume of risk information spanning branches/regions, functional groups and programs/services.
- It is time-consuming and resource intensive to analyze risk data.
- Centralized perspectives of risk intelligence is fundamental to risk management activities.



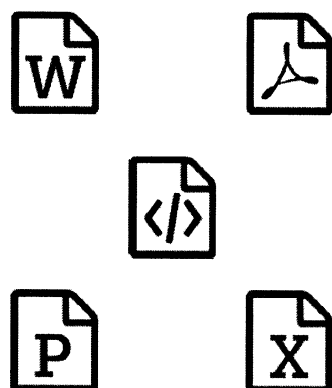
## POTENTIAL SOLUTION: DEEP LEARNING

- Deep learning enables the recognition of patterns in vast volumes of data that would be impossible for humans to process.
- Deep learning can aggregate and analyze risk information from across the Department to develop new views and insights into risk.
- Strong centralized risk intelligence will provide decision support to enable ESDC to identify and respond to risks.

## SELECT DATA

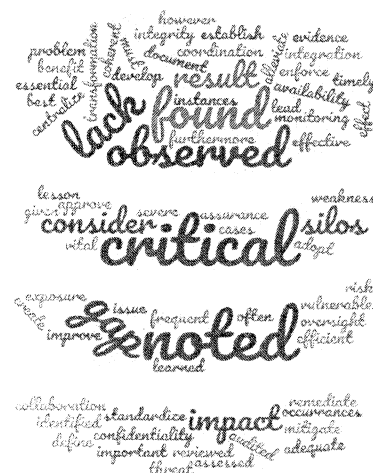
Unstructured data was used in the proof of concept, consisting of audit reports from the past 10 years.

A total of 107 files, 2,154 pages were processed.



## WORD DATABASE

Created a bank of 92 common word embeddings, often seen in audit reports, to signal where important statements would likely be. Words such as: found, however, adequate, inconsistent, critical, etc.



## MACHINE LEARNING

Over 500 statements have been manually categorized as risk or not by multiple audit professionals. This series of binary true/false questions were then fed to the algorithm, which it uses to make decisions on what a risk statement is.

? We also noted that the CIO attends all BPTs meetings to ensure IT is considered in strategic discussions and decisions across the Department

? We also noted that managers and authorized requestors did not consistently apply the least privilege principle and the need to know requirements

✗ We also noted that the CIO attends all BPTs meetings to ensure IT is considered in strategic discussions and decisions across the Department

✓ We also noted that managers and authorized requestors did not consistently apply the least privilege principle and the need to know requirements

## STATEMENT BANK

Machine recognizes patterns of what a likely risk statement is and a bank or list is created. The machine provides a score of how confident it is a given statement is a risk based on the pattern.

99% we found that adding a second reviewer at the in-person centres did not significantly improve the quality of the registration form and added four days on average to processing time.

99% There is a risk that controls over access to sensitive or personal information may become inadequate as AppliWeb evolves if the TRAs and PIAs are not kept current.

98% there is no periodic and comprehensive monitoring of access to detect irregularities.

70% There are various mechanisms used by ISB to identify ineligibility or incorrect payments.

## LINK & TAG

Statements are tagged to the 'ESDC Audit Universe' using: report title; objective; scope; and paragraph. Statements are linked to corresponding auditable entities from the Audit Universe to provide views for analysis.

### Payment Programs

- Benefits to Individuals
- Funding to Organizations
- Transfers to Other Governments

### Services to Citizens & Organizations

- Information Products
- Regulatory, Mediations & Advocacy
- Identity Management
- Service Delivery Channels

### Management Services

- Management Frameworks
- Planning & Accountability
- Asset & Resource Management
- Human Capital Management
- Safeguarding of Assets, Info & People
- Management of Info. Technology
- Communications
- Policy Program Service Continuum

# MODERNIZATION OF THE BUSINESS INTELLIGENCE SYSTEM

*Bilateral and Regional Labour Affairs Division*

July 9, 2019

# Table of Contents

<b>PART I.....</b>	<b>2</b>
i. BACKGROUND	
ii. STRATEGIC GOALS	
iii. SCOPE AND STAGES.....	3
iv. PROOF OF CONCEPT (PoC) FOR STAGES 1 AND 2 .....	4
v. PILOT	
vi. FEDERAL PARTNERS	
 <b>PART II.....</b>	 <b>5</b>
i. STAGE DESCRIPTION	
 <b>PART III.....</b>	 <b>7</b>
ANNEX 1: BASELINE - BRLA's present repository	
ANNEX 2: BASELINE - Repository content, production of intelligence and search methods.....	9
ANNEX 3: BASELINE – BRLA's core activities.....	10
ANNEX 4: Examples of use cases.....	11

## Modernization of the Business Intelligence System of the Bilateral and Regional Labour Affairs (BRLA) Division

### PART I

#### i. BACKGROUND

BRLA has started a process to modernize its Business Intelligence (BI) system with a project to test Artificial Intelligence (AI) technologies for the improvement of information management and analytics. BRLA aims to spearhead innovation for Labour, in line with ESDC's corporate vision and analytics agenda.

Since 2018, BRLA is working collaboratively with ESDC's Innovation, Information and Technology Branch (IITB) to build a strategy to test and implement a technology solution. The units of IITB include Data and Analytics Services (DAS), Business Relations Management (BRM), Business Solutions and Information Management (BS-IMS) and the Artificial Intelligence Centre of Excellence (AI-CoE).

BRLA is also liaising with the Corporate Secretariat (Enterprise Solutions and Security), Strategic Integration and Governance (SIG) and the Chief Data Office (CDO). Externally, the unit is benefitting from excellent liaisons with Veteran's Affairs Canada, Global Affairs Canada and the National Research Council, who have successfully tested and adopted IBM WEX for similar business cases

As a result of internal collaboration, the vendors IBM, OpenText and SAS, who developed comparable AI-technologies, have been identified as possible solution-providers for BRLA's needs. For this, ESDC will first test IBM Watson Explorer (WEX), which is capable of Artificial Intelligence through Machine Learning (ML) applications, and integrates with open-source packages. Subsequently, BRLA will test Viya, the equivalent SAS product.

BRLA has also been liaising with OpenText but, although its platform Magellan has been identified as another possible solution, this vendor has not yet materialized any steps towards a PoC.

#### ii. STRATEGIC GOALS

Improved performance: BRLA aims at leveraging its information management and analysis, in order to enhance business performance.

Innovation: the unit wants to develop new and better ways to work with labour provisions of free trade agreements, and international technical assistance.

BRLA's approach includes productive management of unstructured data, the development of increasingly reliable reporting methodologies, and the implementation of more efficient forms of data collection and analysis.

The unit expects to accomplish innovation with an adequate use of AI-technologies. The main areas expected to report positive changes are:

- Overall productivity
- Decision-making
- Research, analysis and reporting
- Capacity building
- Information Management

*Production of quality business intelligence:* a new system should offer: access to diverse types of data; maximum value extraction from the unit's unstructured information assets; the production of accurate intelligence; and, the capacity of conducting descriptive, predictive, prescriptive and cognitive analytics in an efficient manner.

*Enhanced and growing multi-source repository:* initially, a new system should facilitate access to and management of BRLA's existing repository. Subsequently, it should expand this repository through Labour Program databases and selected external sources such as libraries, EBSCO, university academic material, and data from the International Labour Organization (ILO) and other web-based sources. Finally, the system should access primary data from selected respondents, as well as public social media.

### iii. SCOPE AND STAGES

In July 2018, CIO Peter Littlefield and DG Rakesh Patry received a briefing note from AI-CoE and BRLA with a general description of the project (see Annex 1). The underlying goal was top-down innovation in ESDC by addressing corporative Information Management (IM) issues involving GCDocs.

The project would have started with two use cases, namely hard-drive cleanup and migration to GCDocs, both led by Business Solutions and Information Management Services (BS-IMS). However, in April 2019, BS-IMS confirmed it would not yet be pertinent to prove a concept at this level.

In light of this shift in strategy, BRLA and AI-CoE re-defined the scope of innovation. BRLA and AI-CoE are now teaming-up to achieve bottom-up innovation in ESDC by addressing BRLA's business intelligence needs with Machine Learning (ML) solutions. By doing so, AI-CoE expects to launch an initiative that would involve other directorates within ESDC, and which could potentially benefit from BRLA's experience.

The project considers four stages:

Stage 1: Content analytics.

- From the current unstructured repository, analyze how much information of what kind exists, and where.

Stage 2: Natural Language Processing (NLP) and cognitive analytics for discovery and search

- From the current unstructured repository, the extraction of patterns and insights
- Once the patterns emerge, search for answers to specific questions

- This should include an ontology framework containing the acronyms, lexicon and definitions pertinent to BRLA's business and the federal government.
- This is expected to work, regardless of the functional classification in place

#### Stage 3: Repository's revamp and IM-policy implementation

- Implementation of a functional classification in line with ESDC approved policies and practices, and the unit's logic model.
- Repository maintenance
- Repository expansion to include external sources and social media
- Develop new ways of data collection, at the secondary and primary levels.

#### Stage 4: Exploration of advanced cognitive applications

- Descriptive and predictive analytics
- Exploration of advanced cognitive applications such as Cognitive Trade Assistant (CTA).
- Leverage BRLA's business model to facilitate negotiations, decision-making and technical assistance.
- Artificial Intelligence applications such as Machine Learning

### **iv. PROOF OF CONCEPT (PoC) FOR STAGES 1 AND 2**

BRLA aims at proving that the project's design concept for stages 1 and 2 are feasible with the IBM Watson Explorer (WEX). The primary goal is to demonstrate how the system addresses BRLA's business intelligence needs, and understand what WEX could offer to boost BRLA's productivity.

A demo for DGs and ADMs from Labour and IITB is planned. The demo should prove that the WEX is a reliable and functional tool, suitable for BRLA's present and future needs.

If the PoC succeeds and senior management considers this a viable solution, BRLA would submit an Investment Management Proposal (IMP), in close collaboration with AI-CoE.

### **v. PILOT**

The Innovation and Information Technology Branch (IITB) might opt to start a pilot project after a successful PoC.

### **vi. FEDERAL PARTNERS**

BRLA is in contact with the project leads of Veteran's Affairs Canada (VAC), Global Affairs Canada and the National Research Council of Canada, who have kindly offered their support to ESDC for the present project



These federal departments have already purchased the WEX and are sharing their experience and in cases like VAC, even some material such as functional computer code, which will spare ESDC of a good amount of effort.

## **PART II**

### **i. STAGE DESCRIPTION**

#### **Stage 2: Natural Language Processing (NLP) for discovery and search Analytics and task automation**

BRLA aims at advancing the production of business intelligence (BI), by structuring its information and conducting analytics with the assistance of text mining. With the support of Artificial Intelligence (AI) applications, the unit also wants the automation of resource-intensive tasks such as text summarization, and the search for specific information in a vast, multi-source system.

In this sense, the motive would be to increase productivity by letting the WEX perform simple but tedious, repetitive and time-consuming work currently performed manually, to leave other higher-level duties for the analysts. For example, accurate action research on a current issue related to a free trade agreement such as CUSMA requires the investment of several hours of energy and patience. The more complex the question, the higher the amount of resources required to answer it.

The access to different sources of information, searching for a certain topic, browsing the available resources, reading and selecting material, summarizing results and finally synthesizing them can take an experienced analyst several hours. After indexing and training an ML system, this task should take not more than a few minutes.

Automation should also help with workload. For instance, building automation capabilities should allow a smooth transition from data gathering to information use, in cases where insight, analysis and decision-making is required.

#### **Stage 3: Repository's revamp and IM-policy implementation BRLA repository's revamp**

BRLA is considering revamping its local repository with the help of natural language processing and supervised Machine Learning (ML). The main goal is to attain a manageable system of records for a better extraction of information value. BRLA would implement this model based on ESDC's IM policies.

##### 1a. BRLA's multi-source system

The unit plans to extend its reach of information. This would mean the integration of primary data such as surveys and public aggregate social media, to develop its repository into a multi-source system maintained with the support of ML technologies.

If GCDocs were ready for implementation in ESDC and BRLA had to migrate its repository, it would use the help of supervised Machine Learning (ML) and a model built in cooperation with AI-CoE.

### 3. Analysis from social media

BRLA has been extracting insights from public social media through Watson Analytics (WA), a tool based on IBM-cloud service that allows visualization only. In addition to proving the concept of discovery from massive unstructured sets of information, WA has exemplified sentiment analysis using aggregate data.

WA was recently discontinued; therefore, BRLA will now explore new options of access to social media with tools such as SAS Viya.

#### **Stage 4: Exploration of advanced cognitive applications Cognitive Trade Advisor (CTA)**

Once BRLA has mastered basic analytics, it would start looking at advanced cognitive applications such as cognitive trade advisors (CTA) to assist with decision making through cognitive solutions. In addition to improving the preparation and performance of negotiations, the CTA would complement a variety of cross-sectional analyses.

For more information on Cognitive Assistants, please consult:

<https://www.ibm.com/blogs/insights-on-business/government/rise-cognitive-assistant/>

To read more about CTA, please refer to the following links:

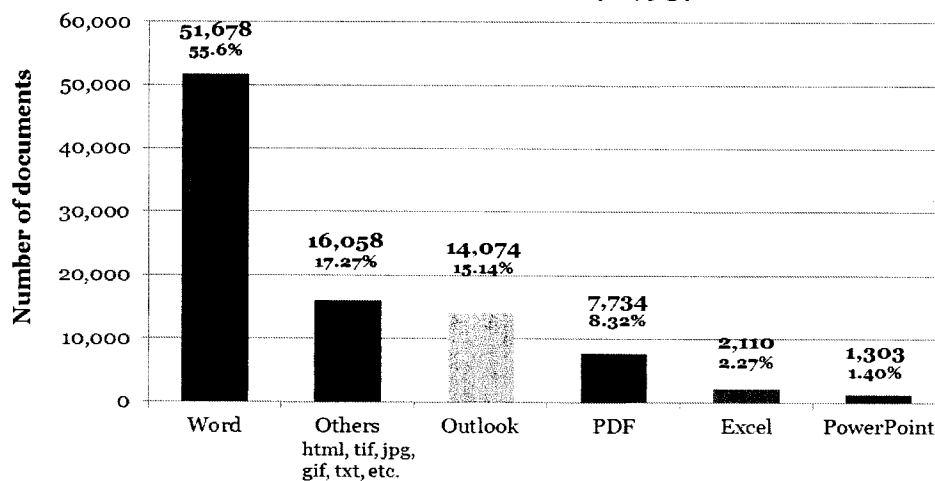
<https://picante.today/latest-news/2018/10/02/6391/icc-launches-artificial-intelligence-tool-for-trade-negotiations/>

<https://bitcoinexchangequote.com/international-chamber-of-commerces-new-blockchain-tools-cognitive-trade-advisor-cta-and-ai-intelligent-tech-trade-initiative-itti/>

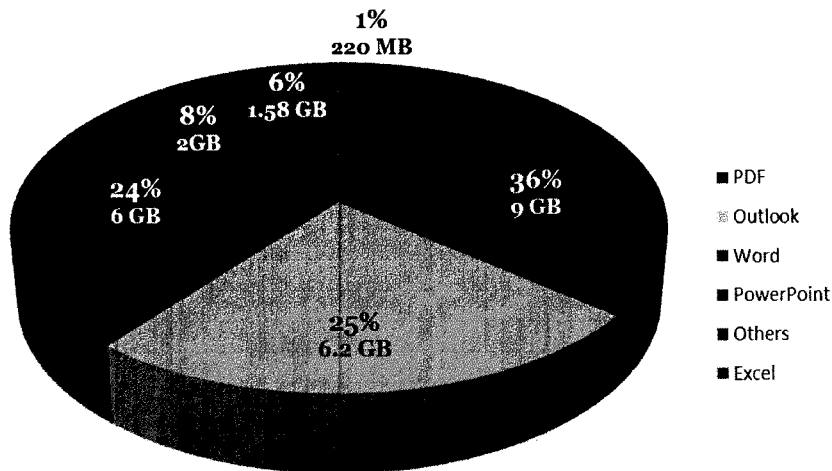
## PART III

### ANNEX 1: BASELINE BRLA's present repository

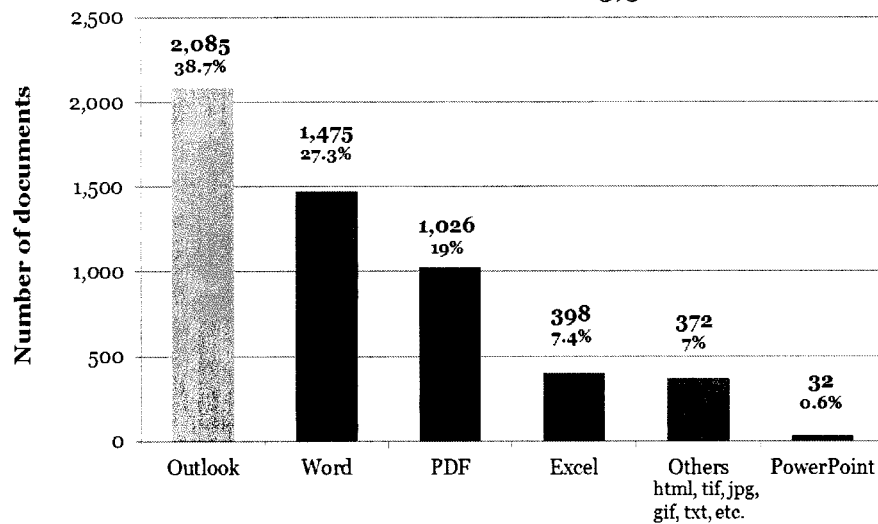
#### U: BRLA estimated total of 92,957 files



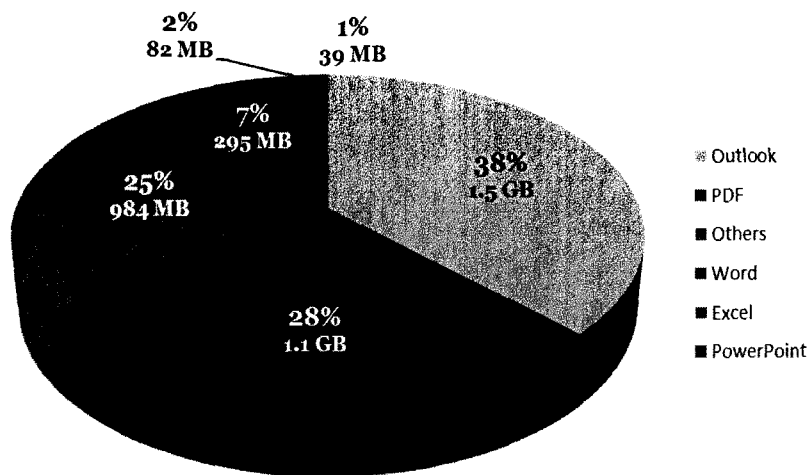
### U: BRLA approximate size 25 GB



### U: LFP estimated total of 5,388 files



## U: LFP approximate size 4 GB



## ANNEX 2: BASELINE Repository content, production of intelligence and search methods

### CONTENT

#### *Agreements and Labour Provisions*

- Documentation on Canadian international trade agreements (TPP, NAFTA, Canada-Honduras FTA, etc.)
- Information on negotiation and implementation
- FTA labour provisions (NAALC, Canada-Colombia and Canada-Chile Agreements on Labour Cooperation, etc.)

#### *Monitoring and Compliance*

- Complaints and conflict resolution, legal and non-legal references
- Reports on countries that have agreements with Canada
- International cooperation activities to support enforcement and compliance

#### *Labour Standards*

- National and International LS, OHS and Industrial Relations
- References from the ILO, Human Rights Commission, United Nations, OECD, US Department of Labour, Global Affairs Canada, etc.)

#### *Promotion of Labour Rights*

- Internationally recognized fundamental rights (collective bargaining, freedom of association, etc.)
- Protection of Canadian workers and employers (economic, legal and informative documents)

#### *International events*

- International conferences and events where Canada participates

#### *Technical assistance and cooperation*

- Technical assistance, including labour funding program agreements (grants and contributions), monitoring and follow-up reports
- Memorandum of Understanding (MOU) and other non-legally binding cooperation

#### *Research and internal notes*

- Documents containing research and analysis of specific topics
- References and publication extracts supporting research (Canadian Parliament Library, online publications and international reports)
- House cards, briefing notes ministerial notes, and backgrounders

#### *Administrative and organization files*

- Operations information, such as the Performance Information Profile (PIP)
- Administrative documents
- Reports of business travel

### PRODUCTION OF BUSINESS INTELLIGENCE

#### *Where does answers to questions/information requests come from?*

- Several questions and issues related to BRLA find an answer with the help of the local U-Drive; however, it is often necessary to consult external sources through conventional Google searches.
- The background and experience of management and senior colleagues add up to the main sources of intelligence. For instance, verbal briefings or hints are useful to start searches in the U-Drive or in other sources such as the Internet.

#### *Manual process to search in the U-Drive*

- If there is the need to search in the U-drive, the process starts with intuitive browsing, until the document(s) possibly containing the required information is (are) identified.
- Once selected, one opens and reads the document(s) in order to assess if the contents are helpful.
- The quest for information normally includes verbal interaction with those who have more knowledge on the U-Drive stored information and the subject matter in question.

### **Annex 3: BASELINE**

## BRLA's Core Activities

Five core Activities	Examples
Negotiation and implementation of labour provisions for Free Trade Agreements (FTAs)	NAFTA/CUSMA, Mercosur, Pacific Alliance, Canada-Peru Agreement on labour cooperation, etc.
Technical Assistance	ILO for Costa Rica, NGOs for Mexico and Vietnam, etc.
Promotion of fundamental labour rights	Freedom of association and collective bargaining, laws against child / forced labour, occupational health and safety, etc.
Participation in international labour forums	Inter-American Conference of Ministers of Labour IACML
Protection of Canadian workers and employers	Level international conditions for Canadian employers and workers

## Annex 4: Examples of use cases

### Stage 1

What kinds of documents are there regarding NAFTA?  
Are there any documents talking about CUSMA?  
Is there any information on technical assistance for Colombia?

### Stage 2

#### Discovery

What insight could one get from the available information on CUSMA?  
What insight can one draw from topic keywords such as U.S., and right to work?  
What insight can one draw from topic keywords such as Peru and labour laws?

#### Search

The classification of documents according to rules such as:  
Freedom of association & Latin America & complaints

1. Complaints against Mexico under the NAALC in the last 10 years?
  - a. Search for documents related to NAALC
  - b. Search with criteria such as Public Communication / complaint
  - c. Narrow down to years equal or greater to 2009
  - d. Narrow down to CAN or US (the prefix for complaints against Mexico)

What are the complaints about?

2. What information exists regarding wages in states with Right to Work?
  - a. Search for documents related to NAALC
  - b. Narrow down to the US
  - c. Narrow down to Right to WorkWhat are the highlights?



NOW AND TOMORROW EXCELLENCE IN EVERYTHING WE DO

# International Labour Affairs (IILA) Bilateral and Regional Labour Affairs (BRLA)

Modernization of the  
Business intelligence  
System

April 26, 2019



Employment and  
Social Development Canada

Emploi et  
Développement social Canada

Canada

## INTERNATIONAL AND INTERGOVERNMENTAL LABOUR AFFAIRS (IILA) BILATERAL AND REGIONAL LABOUR AFFAIRS (BRLA)

IILA's business and the nature of its data, is quite **different from the rest of the LP**.  
IILA's work revolves around highly unstructured data, i.e. narrative and textual types.

Core Activity	Examples
Negotiation and implementation of labour provisions for Free Trade Agreements (FTAs)	Canada-Colombia Agreement on Labour Cooperation, CETA, CPTPP, CUSMA, etc.
Technical Assistance	ILO for Costa Rica, the NGO MSN for Mexico, etc.
Promotion of fundamental labour rights	Freedom of association and collective bargaining, laws against child/forced labour, OHS etc.
Participation in international labour forums	For example, the Inter-American Conference of Ministers of Labour IACML
Protection of Canadian workers and employers	Level playing field for Canadian employers and workers

*Bilateral and Regional Labour Affairs, April 26, 2019*



## Strategic goals:

- ✓ Improved performance: BRLA expects to leverage information management and analysis
- ✓ Innovation: The intelligence produced by BRLA includes **massive unstructured data**, which requires the development of **qualitative methodologies** and cost-efficient forms of data collection and analysis. BRLA expects to achieve innovation with an adequate use of **text analytics, natural language processing and AI-applications such as Machine Learning**.
- ✓ Production of quality business intelligence: Optimal access to diverse types of data; **maximum value extraction from BRLA's unstructured information assets**; and the capability of conducting **descriptive, predictive, prescriptive and cognitive analytics** in an efficient manner.
- ✓ Enhanced and growing multi-source repository: Facilitate the access to and management of the existing repository. Subsequently, expand this repository, by reaching out to databases within the Labour Program, as well as external and web-based sources (e.g. Statistics Canada, Global Affairs Canada, EBSCO, International Labour Organization, etc.). Finally, access to **primary data from selected respondents and public social media**.

*Bilateral and Regional Labour Affairs, April 26, 2019*



Employment and  
Social Development Canada

Emploi et  
Développement social Canada

Canada

### Steps already taken:

- Our unit started by assessing the Collective Agreement Search Application (CASA), conceived and developed by the Chief Data Office, in order to explore a possible in-house solution for BRLA's business needs. However, the CASA could not prove its concept for any use case.
- For over 9 months, BRLA has been successfully using text analytics from public social media to complement the production of business intelligence.
- The unit is currently assessing Watson Explorer (WEX). This system has **Machine Learning (ML) capabilities and integrates with open-source packages.**
- To ensure IILA develops a solution in line with ESDC's corporative strategy, we have been **working collaboratively** with ESDC's Innovation, Information and Technology Branch (IITB), including Data and Analytics Services (DAS), Business Relations Management (BRM), Business Solutions and Information Management (BS-IMS) and the Artificial Intelligence Centre of Excellence (AI-CoE).
- In addition, BRLA is liaising with the Corporate Secretariat (Enterprise Solutions and Security), Strategic Integration and Governance (SIG) and the Chief Data Office (CDO).

*Bilateral and Regional Labour Affairs, April 26, 2019*



**BRLA and its partners have planned four stages :**

**Stage 1:** Mapping existing intelligence through **content analytics**.

**Stage 2:** Obtaining insight (**discovery**) and finding relevant information (**search**) using natural language processing, ML and other text analytic features.

**Stage 3:** Current information repository's revamp (**IM**) and integration with **social media**.

**Stage 4:** Exploration of **cognitive** applications

- BRLA is **engaged in discussions and consultations with federal government departments**, who have successfully tested and adopted IBM WEX for similar business cases as BRLA.
- DGOs and ADMOs from both BRLA and AI-CoE will have a **demo to prove the WEX concept (PoC) with its use cases**, in order to assess stages one and two.

*Bilateral and Regional Labour Affairs, April 26, 2019*



**Moving forward :**

- ❖ IILA will continue to spearhead innovation, by planning the modernization of its Business Intelligence (BI) system, through the improvement of its information management and analytics with Artificial Intelligence (AI) technologies. This is an **ongoing work with IILA's internal and external partners.**
- ❖ BRLA recently hired another student to support this process.

*Bilateral and Regional Labour Affairs, April 26, 2019*



Employment and  
Social Development Canada

Emploi et  
Développement social Canada

Canada



Employment and  
Social Development Canada

Emploi et  
Développement social Canada

Canada



# Chatbot for Tier 1, Passport General Enquiries Service

## Concept Test: Final Report

Service Transformation Committee (STC)

June 18, 2019

1

# Overview

Project objectives	Key findings
Assess how chatbots can leverage 1-800 O-Canada's general enquiries service – using Passport program as test case	<ul style="list-style-type: none"> <li>• 1-800 content needs revision to support chat interactions</li> <li>• Tier 1, general services, seem well-suited to chat services</li> </ul>
Inform ESDC's STP and PMT activities	<ul style="list-style-type: none"> <li>• Content must be structured</li> <li>• Content style must be conversational, two-line snippets</li> </ul>
Inform future channel shifts as per ICMS	<ul style="list-style-type: none"> <li>• Long term investment in staff skills to manage high value content</li> <li>• Information architecture tools are needed (drag and drop)</li> </ul>
Assess needed technology/web requirements and internal business processes for chatbot services	<ul style="list-style-type: none"> <li>• Business requirements and web specification available</li> </ul>



# The Test

- Live test on main Passport page of Canada.ca
  - March 18<sup>th</sup> to April 5<sup>th</sup>
  - not promoted on other passport pages
- Initial “soft launch”
  - chat link offered at the bottom of the page for an hour in the morning and again in the afternoon
- By end of test - chat link offered from 9 a.m. to 3 p.m. at the top of the page
- CSB staff monitored chatbot performance, stepping in with corrections or new content as needed
- Steady demand – avg. page visits were less than the 800/hour estimate, with uptake of the chat under 3%
- Bot performance improved with use - making its answers more reliable over time

## Week 1 - Operator-assisted mode test

- All bot messages required operator approval
- Bot availability initially set to 10 - 11:30 a.m. and 1 - 2:30 p.m.
- Based on bot performance, availability was slowly extended throughout the week

## Week 2 - Operator-assisted and Automatic mode test

- Bot availability was expanded to cover 9 a.m. – 3 p.m.
- Half-way through the week, bot was set to Automatic mode allowing it to answer questions on its own if it was able to achieve an 80% confidence threshold. If not, the bot would escalate to a live Operator to take over client session
- Every client session still required Operator oversight and observation

## Week 3 - Automatic mode test

- Bot availability remained at 9 a.m. – 3 p.m.
- Plan was to move the chat invitation to the top of the Passport Services page on the Canada Site. However publishing issues delayed that until April 4<sup>th</sup>

## April 4 & 5 - Invite was moved to the top of the Passport Services page on the Canada Site

- Bot traffic more than doubled on those two days based on the previous average of the preceding days.

# Key Insights...

- Content style and structure is foundational to support a chat service:
  - Over the course of the pilot, significantly more time was spent on revising existing service content to meet the requirements of a chat interaction than configuring the technology and conducting the test...
- Usability of the chat interaction was improved by anticipating and developing follow-up questions with “Quick Reply” standardized answers
  - allowing clients to avoid personalization and simply choose a prepared next question
- Address potential bias at the content level
  - ensuring diversity in the make-up of teams working on content and data rules
- Staff performing the operator role were able to manage about 3-5 chat interactions concurrently.
  - 8 operators may be needed to support a full service for a busy program, but a “live operator” determination model needs more testing
- Accessibility and usability of the chat need special attention to ensure success
  - Some overlap in query and response between users and the bot

# The Passport test vs other chat technologies

*Need to choose the right model in a fast moving technology environment*

Many GC services are adopting chatbots:

- CRA tested a chatbot this tax season
- ESDC's National Student Loans Service Centre runs an FAQ navigator / chatbot
- ESDC's LMI Explore launched a NOC chatbot (Nick)

"Ask CRA?"  
"Ask Nick?" } Chatbots

More powerful chats combine chatbot + Natural Language Learning, Understanding (NLU) and Generating, to understand the "meaning of words"

"Ask Siri?"  
"Ask Alexa?"  
"Ask Google?" } Voice Chat

## **Passport Chat tested a hybrid chat service:**

- Automatic and Operator-Assisted modes
- Supports organizational change management
- Hybrid, open source tech is content agnostic
- Uses instant messaging (IM)

Passport's  
"Ask SC Chat" } Hybrid Chat Service

Hybrids can add UX and enhance the chat experience with related content



# What our Hybrid Chat Service included...

*Distributed nature of hybrid, open source chat technology*

Specialist chatbot focused on individual service:

- Controls the initial reply
- Provides dialogue management tools
- Identifies client intentions from existing content

Master NLU focused on learning new content:

- Controls the classification of service content to individual specialist chatbots
- Machine learning and training capacities
- Identifies new client intentions, adds content

Added UX:

- Controls the presentation elements of the page, related content, Quick Reply buttons
- Can control the URL of the browser
- Available API connectors for Facebook, Twitter, IM, etc.
- Third-party add-ons available, open source

same platform

web presentation

Hybrid Chat Service escalated clients to a human operator – the most trusted source of info

Open source tech with some proprietary algorithm for analyses

GC owns the content which can be easily “lift and shift”

# After the test, CSB sought Industry expert advice...

*Industry standard in adopting chatbots is to focus on optimizing service content*

- Off-the-shelf enterprise chatbots (i.e. “shrink wrapped” with Adobe, MS or other platforms) need significant customization to develop a hybrid service – effort should be on staff and content, not technology
- To support a chat interaction, content managers should focus on improving content into a conversational style with parent-child relationships between data points
- Call centre staff mostly interact with the chatbot in an operational setting, integrating chat service with telephony, email, etc.
- Typically, service content that supports chat interactions benefits other service channels as well

*Choose “best in breed” technologies to support staff and avoid major technology investments in this fast paced environment*

*Sources – Gartner interviews and literature review*

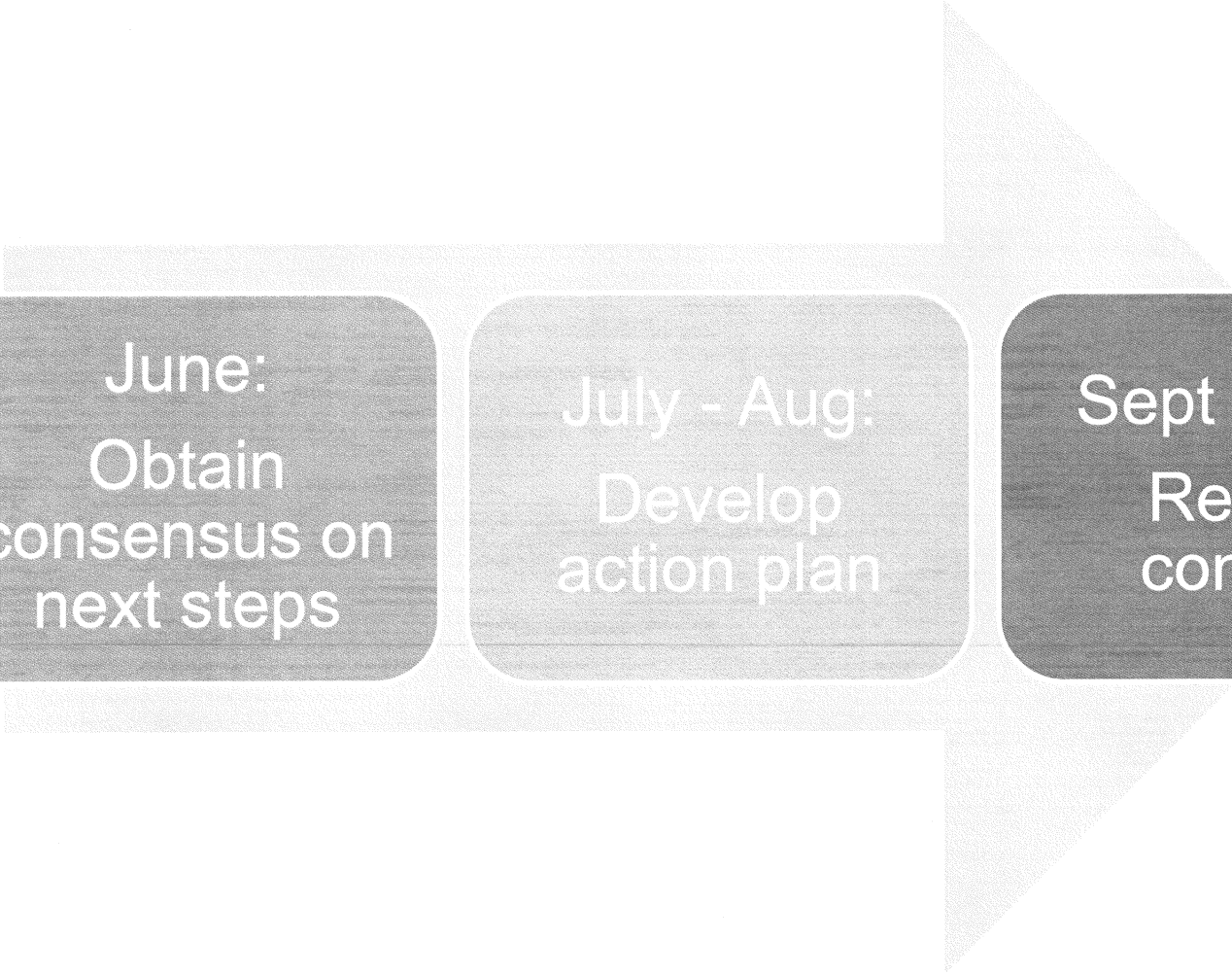


# Next steps

Based on these test results, CSB can:

- Extend the current Passport chat for one year, as ESDC considers a longer term approach to chat bot services.
  - Testing a second program over the coming year
  - Refining the model to re-skill call centre staff to take on new chat services – resource model could be determined with more testing
  - Further improving the user experience (explore new features, related content)
- Contribute to a broader enterprise chat service approach
- As part of web optimization and voice search optimization efforts - Reuse chat content on Canada.ca where it can be leveraged by other voice technologies like Google Assistant, Siri, Alexa
- Support ESDC's ethics, anti-bias efforts by leveraging gender and ethnic diversity in the make up of teams working on content and data rules

**Leverage the chat tools and framework to help optimize service delivery**



June:  
Obtain  
consensus on  
next steps

July - Aug:  
Develop  
action plan

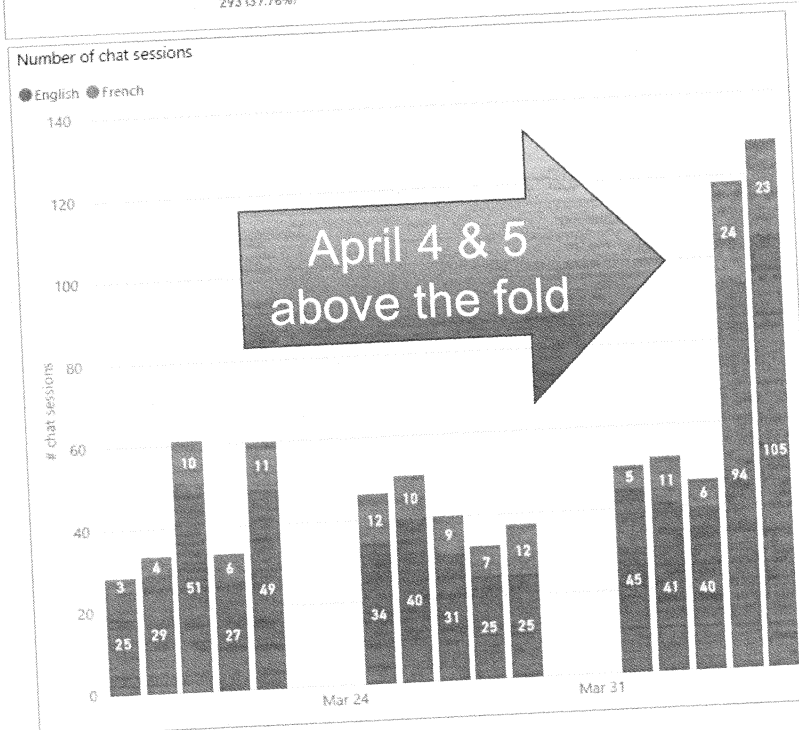
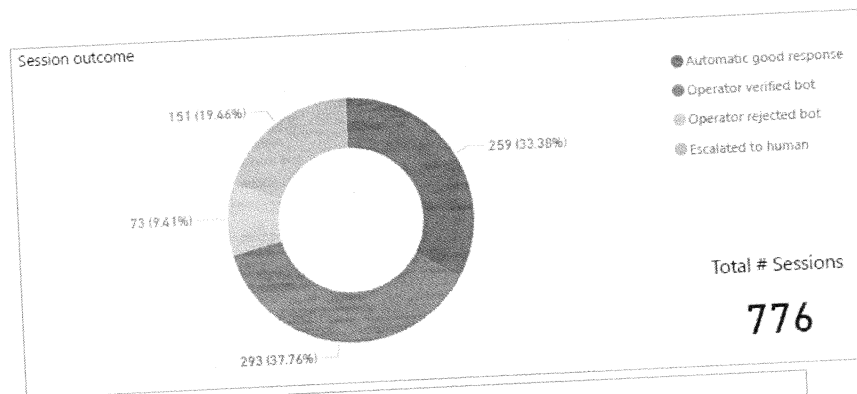
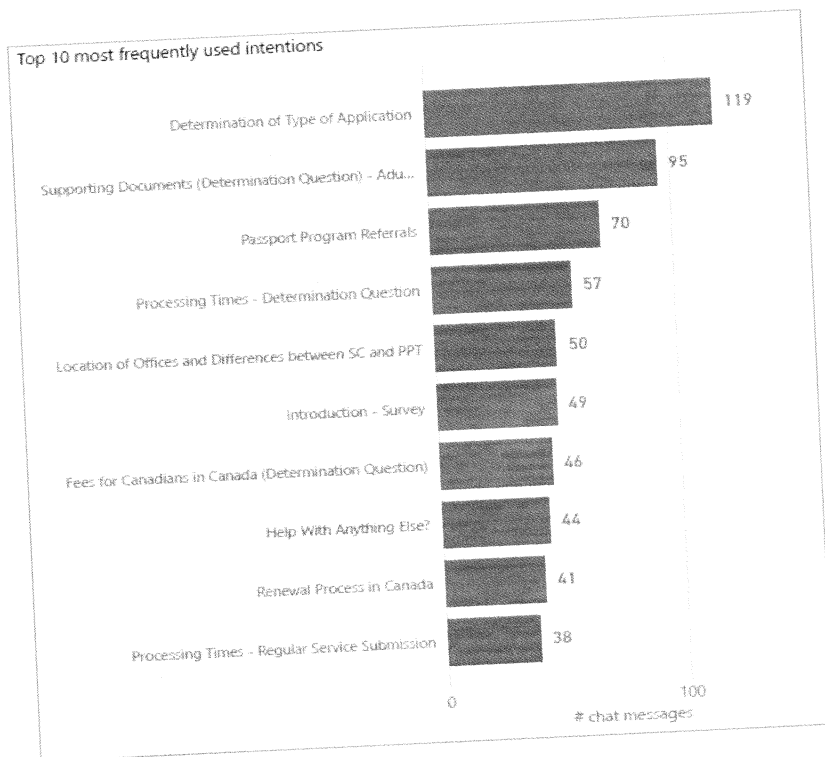
Sept – Mar:  
Revise  
content



# Annex: Test Results

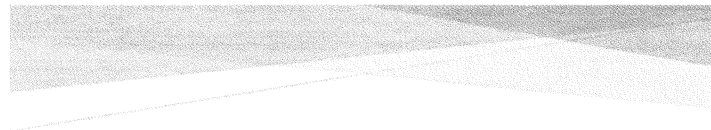
Live testing performance of chat service:

- 776 client sessions; 2134 client questions
- 769 client clicks of standardized answers
- 71% of bot answers were correct without needing operator interventions





# Annex: Evaluation document



## SC CHAT – CONCEPT TEST

Evaluation

### Abstract

Lessons learned and recommended next steps for Service Canada Chat  
Services

Digital Services Directorate  
Citizen Services Branch

Contact Portfolio Web for a copy of  
the evaluation document



# SC CHAT – CONCEPT TEST

## Evaluation Report

### Abstract

Lessons learned and recommended next steps for Service Canada Chat Services

Digital Services Directorate

Citizen Services Branch

## Service Canada Chat – Evaluation of Concept Test

### Getting ready for Intelligent Chat

Modernization of service delivery is leading to an explosion of chat technologies. From the simple, FAQ-supported chatbot to more complex, hybrid chat services, the industry is rich in open source platforms vying to produce the best customer experience.

- ✦ Chatbot technology – simple content, requires exact matches question-to-answer
- ✦ Natural Language Learning/Understanding/Generating (NLG) – seeks to understand the meaning of words, capable of handling inexact match question-to-answer.
- ✦ Hybrid Chat Service – provides additional context to the chat, goes beyond Q/A

Most of these technologies use open source platforms, with some proprietary logic or algorithm to the machine analyses. However while the technologies are ubiquitous and thick on the ground, the content is still the main challenge to be solved in chat interactions.

### Content is a principle requirement for the Chat space

Irrespective of the platform or technology, the performance of content is critical within the web channel of service delivery. The same style of service content provided to Canadians on the phone should also work for machine assistants like Siri, Alexa, Google Assistant... and chat services benefit by the same style and format.

- ✦ The style of existing service content must become conversational, two-line snippets of information – tested by reading out-loud
- ✦ Available service content should relate directly with service delivery requests
- ✦ Web document structure must be unique and provide context to the information

### Testing 1 800 O Canada service content in the Chat space

Service Canada tested the service delivery knowledge repository used by 1 800 O Canada call centres to see if it can support a hybrid chat service and more complex client interactions.

- ✦ The solution should support automatic chatbot features
- ✦ The solution should support machine learning new ways of asking the same question
- ✦ The solution should be capable of providing an enhanced and interactive client experience across media devices
- ✦ Content curation tools should be available to optimize and export key content to other areas of the web channel that could benefit from optimized service delivery content

## Purpose/Objective:

### Purpose of the test

The purpose of testing was to explore business feasibility and technical requirements for implementing chatbot services in the Tier 1 service delivery context. The test aligns with ESDC's overarching CRAWL, WALK and RUN activities for Human Language Search and Dynamic FAQ delivery.

### Needs and Gaps

Need for modernization of service delivery, more self-serve options: Nothing saves time and promotes satisfaction than the ability to ask a quick question... and get almost instant answers about GC services.

Chatbot technology offers business the chance to respond to a quick question, putting program application information in the hands of citizens in a timely fashion.

### New tools and processes tested

A chatbot system that has capacity to learn answers to questions was tested, with ground covered on content structure requirements, content association tools, service protocols in the automated space, accessibility challenges – new business processes and procedures were explored.

- ✦ Call Centre Robot technology developed by Korah Ltd. was tested as part of the Build in Canada Innovation program (BCIP) at PSPC; ESDC is a designated testing department for this technology
- ✦ Web standards for Chat were tested for accessibility, usability and interoperability
- ✦ Service content optimization procedures were tested at the Integrated Channel Management (ICM) back-end office that manages the Information Management System (IMS) at Service Canada. (The IMS supports the IMPACT database used by Call Centre staff)
- ✦ Client up-take and stakeholder feedback mechanisms were tested in a live service setting

## Scope:

### Scope of the testing

The scope of content was limited to the Information Management System (IMS) at Service Canada. The scope of the service was limited to domestic passports and citizens residing in Canada.

### Timeframe

Live testing ran during business hours, Eastern Daylight Time, from March 18 to April 5, 2019, on the Canada.ca Canadian passports theme page.

### Deliverables

The deliverable for the test is this evaluation report, with detail of challenges encountered, solutions and processes used and practical recommendations for next stage development.

- ✚ Channel shift intelligence
- ✚ Business requirements and web standards
- ✚ Optimization requirements for 1 800 O Canada knowledge base
- ✚ Stakeholder feedback
- ✚ Cost and value
- ✚ Recommendations for next steps...



## Stakeholders, Roles and Responsibilities:

### Stakeholders

#### *Internal stakeholders*

- Integrated Channel Management (ICM)
- SC/ Passport Call Centres
- Portfolio Web

#### *Partners and observers*

- Branch partners at STP and CDO
- Public Affairs and Stakeholder Relations Branch (PASRB)
- Immigration, Refugees and Citizenship Canada (IRCC)
- Build in Canada program at PSPC

#### *Private sector vendors*

- Korah Ltd.

#### *External stakeholders*

- Local, in-Canada users of domestic passport services
- General public

### Stakeholder Needs

- Integrated Channel Management (ICM):
  - Test whether Chat service could utilize existing knowledge repositories
  - Test the readiness of business resources associated with Chat services
- SC/ Passport Call Centre:
  - Review chat content against telephony Q & A
- Portfolio Web:
  - Test the web presentation of Chat services
  - Test boiler plate privacy and disclaimer texts
  - Test accessibility code and web standards for Chat services
- Organizational:
  - Inform future STP solutions and PMT projects
  - Inform next steps in channel shift to AI tools, per ICMS
  - Assess how Chatbots can complement and leverage 1-800 O Canada general service
- Korah Ltd.
  - Test its chat technology against GC requirements
- General public, local users of domestic passport services
  - Save time by getting quick answers online

## Stakeholder Roles

- Integrated Channel Management (ICM) stakeholders:  
Represent the interests of the end users of the solution, and help define and validate the content requirements and systems design. ICM leads in User Acceptance Testing and signs off on the usability and accuracy of Chat features and tools.
- Portfolio Web stakeholders:  
Represent the interests of the business area that is sponsoring the test, and provide project management and oversight to coordinate roles, responsibilities and activities for the test; track business requirements. PW leads the technical implementation of Chat on the Canada.ca website and provides expert guidance on accessibility and web standards.
- SC/ Passport Call Centres:  
Provide feedback to ICM stakeholders in optimizing content used within the Chat service.
- Korah Ltd.:  
Provide configuration that conforms to business specification and web standards. Korah is responsible for the disposition of chat service and working within PSPC hosting arrangements.
- These roles/responsibilities persisted throughout the lifecycle of the test and the teams worked well together, gaining in knowledge and expertise of the Chat space.

## RESULTS:

### Calibrated to build trust

Service delivery is all about trust. To build trust and brand recognition for Service Canada Chat, Integrated Channel Management (ICM) took a calculated, low-risk soft launch approach. Since ICM had only a small team of available Operators, it also needed to ensure confidence on the usability of the chat interface before allowing heavier traffic. The goal was to make sure that every single client who clicked on the invite to participate would have the opportunity to do so without being met with any queue or wait times.

#### Week 1 - Operator-assisted mode test

- ✦ All bot messages required operator approval
- ✦ Bot availability initially set to 10 - 11:30 a.m. and 1 - 2:30 p.m.
- ✦ Based on bot performance, availability was slowly extended throughout the week

#### Week 2 - Operator-assisted and Automatic mode test

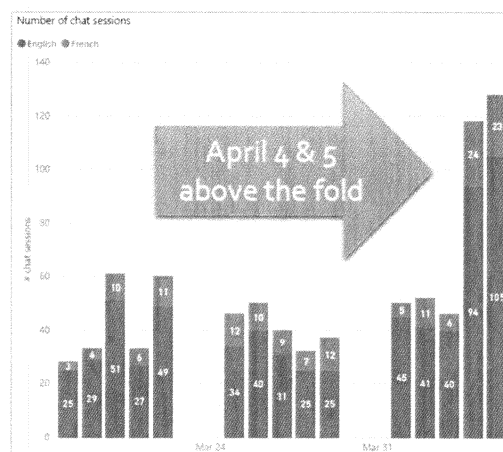
- ✦ Bot availability expanded to cover 9 a.m. – 3 p.m.
- ✦ Halfway through the week, bot was set to Automatic mode allowing it to answer questions on its own if it was able to achieve an 80% confidence threshold. If not, the bot would escalate to a live Operator to take over client session
- ✦ Every client session still required Operator oversight and observation

#### Week 3 - Automatic mode test

- ✦ Bot availability remained at 9 a.m. – 3 p.m.
- ✦ Plan was to move the chat invitation to the top of the Passport Services page on the Canada Site. However publishing issues delayed that until April 4th

#### April 4 & 5 - Invite moved to the top of the Passport Services page on the Canada Site

- ✦ Bot traffic more than doubled on those two days based on the previous average of the preceding days.





## Channel Shift Intelligence

Insight of the strategic considerations to adding a new sub-channel for chat services

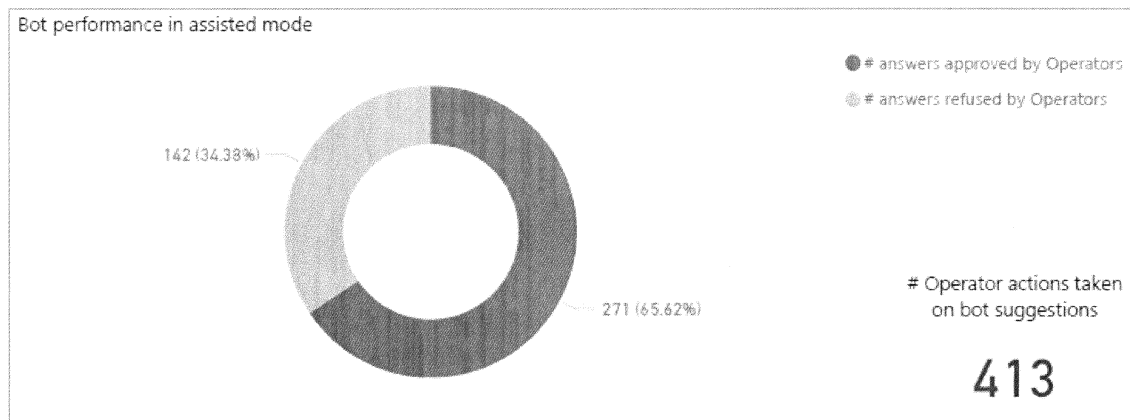
### Automatic Chat performance

Live testing started in operator-assisted mode, and as confidence in the machine responses improved, the bot performed in automatic mode to close out the testing. The more learning sessions the Chatbot experiences, the more reliable and accurate the automatic responses become. On average, a machine confidence level of <80% was used to trigger escalation to a human. In total, there were 2134 questions received over the course of three weeks of testing.

- ✦ Offer rate of the chat to clients was 100% - the chat was available for an hour in the morning and afternoon to start, by the half-way mark of testing it was almost continuously available during office hours (9AM to 3PM EDT): March 18 to April 5
- ✦ Uptake was increased significantly by moving the chat to the top of the passport service page
- ✦ On average, 800 people/hour visited the page and participated in 100 sessions (placement at top)

### Chatbot in operator-assisted learning mode:

A human operator is available to verify each and every answer of the Chatbot. Humans also help improve machine learning by curating service content in the system – associating variable questions with good answers, including verifying machine responses in real time when escalations occur.



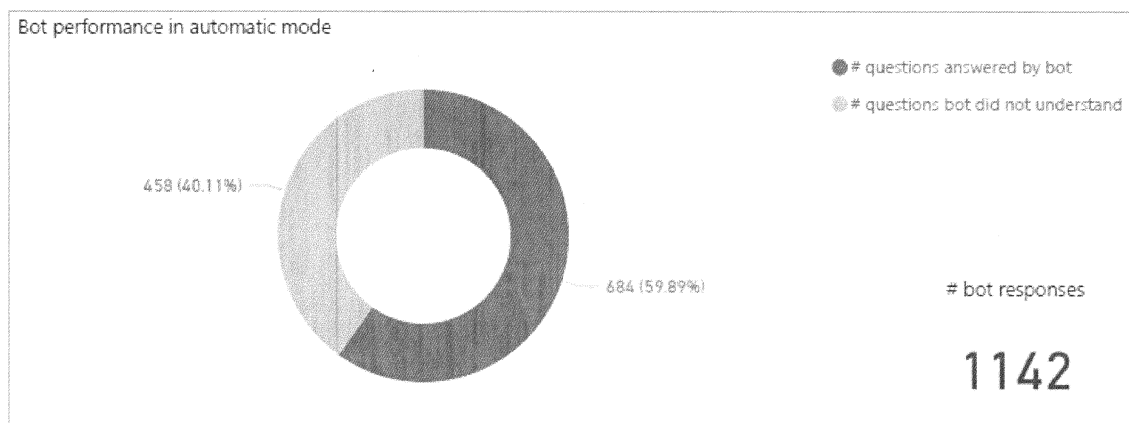
### Considerations

Generally, the human operator takes a few minutes to match a new answer so there is a potential of clients waiting for an operator when an escalation occurs. **Queue times** during escalation were not significant. Mitigation: Clients would be informed of their place in queue, and later if the wait went over a few minutes, the client would be informed that the system is busy, to try again later.

- ✦ Operator to client ratios range from 1:3 to 1:5
- ✦ Human resource capacity during testing was sufficient at 3-4 humans for each day of testing

### Chatbot in automatic mode:

The vast majority of questions fall within the defined mandate of the pilot, i.e. Tier 1 simple questions related to passport services. By mid-way through week 2, the machine was performing well enough to switch to automatic mode: the machine tries twice to answer a client question, but if it fails to reach its 80% confidence level, it escalates to a live operator to take over the conversation. The operator then manually searches for the correct answer and associates it to the question. Seventy-one per cent of bot answers were correct without needing operator interventions.



### Considerations

In an automated setting, even with human monitoring, as was done during testing, there is limited capacity to respond to threatening messages since the clients are unknown. Existing call centre protocols do not necessarily apply when there is no one there to monitor the chat service.

- ✦ During automatic mode, operators were on standby for any issues that arise
- ✦ If there was no human available, machine response to questions it cannot answer without human help is to direct users to call the 1-800 O Canada phone number during office hours
- ✦ Some clients were not interested in answers and simply asked rapid fire questions on different topics – the machine performed well in these situations, as it was designed to do
- ✦ The Chatbot is set-up to strip bad language and SIN from its content archives
- ✦ Machine response to *mauvais mots* was to ask the client to please rephrase their question – content was also added to remind clients of the mandate and purpose of the chat
- ✦ For personal information clients were referred to the Information Notice that directs them not to share personal information in the chat window
- ✦ No threats were received
- ✦ **Machine response to threat messages in automated chat space** – this was a consideration but the issue did not arise during testing. Mitigation: Service Canada would direct users to external resources, i.e., TBS example “If you are in distress, please contact your nearest distress centre. If it is an emergency, call 9-1-1 or go to your local emergency department.” The system could also be configured to send an alert to a specified email when such a message is received. This scenario is only valid in automatic mode without humans.

## Quick Reply Buttons

Based on staff research, Service Canada anticipated possible follow-up questions and created answers that removed the need for them. Quick reply buttons with standardized answers helped provide clients with all relevant content on certain questions. Early indications are that the quick reply buttons are useful in managing Tier 1 content, allowing clients to self-serve and avoid personalization of their questions.

Quick Reply Button	# times used
How do I apply or renew a passport?	81
No, that's all for now.	54
What are the supporting documents required with my application?	39
I am outside of Canada.	37
Supporting documents for an adult.	36
What are the processing times for an application?	35
Yes, I have time to spare.	31
Yes, I have additional questions.	29
Comment faire une nouvelle demande ou renouveler un passeport?	26
I need to apply for a new adult passport.	25

## Operational Considerations

### Integrated Channel Management (ICM) team

For the testing phase, the ICM team played a dual role in revising content for the chat environment and supporting the bot during operations whenever an escalation occurs. In a steady state, the ICM team would not be the ones supporting the bot during operations, it would be call centre agents.

- ✚ Content for the chat environment was revised and structured by ICM team
- ✚ Bot was also assisted by the same ICM team

### Call Centre Agents

Extended testing would help answer the outstanding operational questions about integrating chat to the Call Centre as the pilot was not long enough to offer any insight on these important aspects.

- ✚ Dedicated call centre agents replacing the ICM team during operations to handle bot escalations; or
- ✚ Integrated call centre agents handling multi-queued chat/phone/email interactions

### Resource Determination Model

Feasibility of operational call centre support for chat bot implementation needs further testing.

- ✚ Agent capacity : ability to handle multi-queued chat/phone/email interactions
- ✚ Client usage patterns : best hours of operation for a chat bot
- ✚ Bot learning curve impact on the resource determination model for call centre support
- ✚ Ability to use the chat bot after regular business hours in fully automatic mode

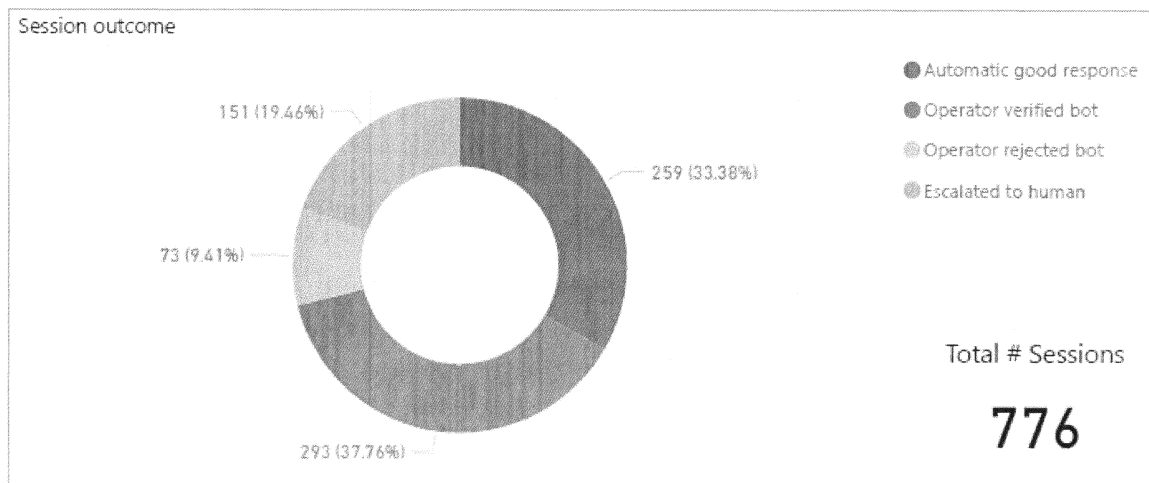
## Optimization requirements for 1 800 O Canada knowledge base

### Suitability of knowledge repository

Content structure and conversation style is very relevant to the chat space. Even though the Integrated Channel Management (ICM) team worked from an established database supporting the 1 800 O Canada repository, the content still required a lot of work to reformat in order to support a chat service delivery model. The content of the answers was already approved which helped, but reformatting so that it could be used effectively in a chat environment is necessary and time consuming.

### Optimized content drives better session outcomes:

The content was developed, mapped and tested using a mix of operator-assisted learning and automatic modes to help the chatbot learn the content associations for passport services.



### Considerations

- ✦ Adapting the knowledge base to support the Chatbot was achieved by training the existing Integrated Channel Management (ICM) team to revise content style, add structure
- ✦ A copy of passport content was supplied to the Chat environment for written revisions and structure was applied using content association tools provided for the purpose
- ✦ Training on the new tools was not difficult and the ICM team obtained tool enhancements during User Acceptance Testing (UAT) – WYSIWYG editors, Split screen FRA/ENG, etc. details have been captured as business requirements in the chat space
- ✦ As new content is added to the 1 800 O Canada knowledge repository, an update is provided to Chatbot – both scheduled and ad hoc updates are supported
- ✦ A non-disclosure agreement (NDA) is in place to protect the 1 800 O Canada knowledge repository from infringement

## Stakeholder Feedback

### Internal stakeholders

Staff reaction has been mostly positive. Regional staff have shown an interest in testing out the chat service for themselves and have offered content suggestions for improvement.

### Partners at STP, CDO and IRCC

Departmental colleagues have participated in brain storming various business requirements. Best practices for terms of use, privacy and accessibility are promised to be shared. Passport operators and partners at IRCC reviewed /signed off on passport content – IRCC provided written feedback on the content and presentation of the chatbot.

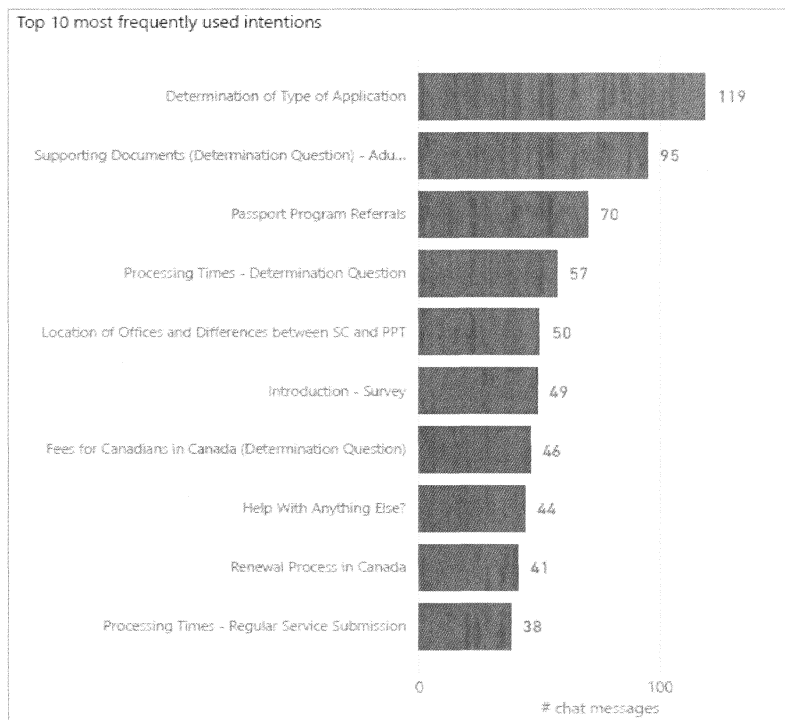
- ✚ The Chatbot experienced a spike in traffic as internal staff and partners tried to trip it up with rapid, unrelated questions
- ✚ Negative testing did not crash the Chatbot

### External stakeholders

Clients reacted positively – out of an average 800 visits/hour to the web page, 100 used the chat service at its peak, uptake was under 3%. The behaviour observed was that clients dropped in to ask a quick question and left. Additional testing time is needed to gauge if the chat service can affect other Tier 1 channels, telephony and in-person.

### French and English Visitors

- ✚ 776 client sessions, 2134 client questions, 769 client clicks of standardized answers



## Business Requirements

### Content association tools

Information architecture (content in parent-child relationships) and content revision was a key requirement to optimizing the 1 800 O Canada knowledge repository to support a chat environment.

- ✚ WYSIWYG editor useful for operators to manage HTML content in the system
- ✚ Free form access needed to revise content or create new content on-the-fly
- ✚ Dialogue management, FRA / ENG co-located content edit screens
- ✚ Drag and drop functionality for content mapping, to quickly make new associations
- ✚ KPI dashboard reporting, export tools to share and review service content collections
- ✚ Scheduled and ad hoc process to update chat environments

### Web Standards

Usability (UX) and accessibility testing was conducted and the final report is being worked into web standards for the chat space.

- ✚ There are two areas to interact with the chat conversation, the query and the response, and this caused a dilemma on where focus should be following an interaction. To create consistency across interactions the focus returns to the query after each interaction. With the response area identified as an "ARIA live" region, every time a response comes in, a screen reader would read the new content out-loud, while the focus is on the query, ready for the next question.
- ✚ The accessibility code fixes implemented by Portfolio Web were reviewed by an accessibility expert and praised as "the best implementation of ARIA live" that they had seen

There are outstanding accessibility issues related to how individual browsers interact with various screen readers, additional steps are required to assure accessibility for all screen readers.

- ✚ Use of iFrame to launch the chat button can affect document structure for screen readers
- ✚ Screen readers' handling of chat "focus" varies by browser, clients may miss stacked messages
- ✚ Additional document structure is needed, headings within mark-up – affects skip links
- ✚ Web specification for chat is available, including details of coding best practices and UX layout

Boilerplate text developed during the course of this test may be reused in other chat implementations, with minor edits.

- ✚ Terms of use
- ✚ Information notice, privacy

Cross-domain issues were encountered and addressed using a sub-domains strategy

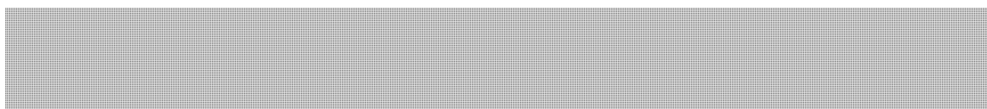
- ✚ Sub-domains have been set-up for chat on the Canada.ca root domain
- ✚ Naming convention allows for multiple chat implementations
- ✚ service1.chat.canada.ca / service1.clavardage.canada.ca

## Cost & Value

### Staffing

Existing staff on the Integrated Channel Management (ICM) team trained to support the chatbot. Due to the length of the pilot and the limited number of client sessions, ICM was not able to determine what would be a proper resource determination model to support ongoing Operator activities. Queue service standard established for the pilot was not tested and ICM was not able to determine a proper learning coefficient for onboarding of new Operators. Further testing and onboarding of additional content is required to determine the number of resources needed to support content creation and maintenance, as well as the number of Operators needed to support chatbot client sessions.

Level of effort for the pilot included:

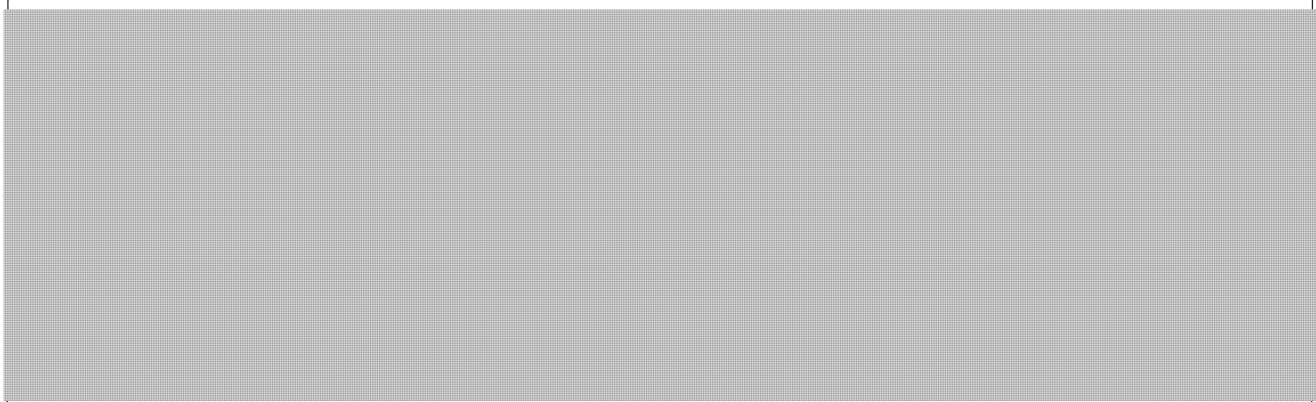


One Web resource is required to monitor the chat service and trouble shoot the queue



Call Centre Robot by Korah Ltd.

ESDC is the principal testing department for Call Centre Robot, through the Build in Canada Innovation program (BCIP) at PSPC. As such, the technology was available to Service Canada for testing purposes without cost. Costing for using the technology beyond the BCIP timeframes are as follows.



### Demand for the Service

Modernized service delivery is on the wish list of most Canadians broadly speaking. During testing, the chat service was soft-launched at the bottom of the page – when it was moved above the fold to the top of the page, the number of transactions doubled. A well-promoted, hard launch is expected to increase volumes dramatically. Placed at the top of the page the chat service uptake was roughly 3% of the 800 visitors to the page each hour.

- ✦ Demand for this service is steady throughout the year
- ✦ Ability to ask a quick question is relevant to this service

## Value

### Scalable approach

Creating service delivery content to support AI is a good idea in general. AI are largely used as an assistant and are not independently intelligent. For example, in order to use AI within the service delivery context, content must be revised for reading out-loud. There are two types of AI that are industry standard.

- ✚ Symbolists – AI that uses logical rules and representations (SC Chatbot fits here)
- ✚ Connectionists – AI that relies on biology inspired neural networks (Watson, Clara are examples)
- ✚ The future of AI is a combination of the two approaches, so that each is needful

Change management for AI should leverage the symbolist approach in the early stages. Once the foundation of content rules and associations is available, the connectionist approach will work better.

- ✚ AI is driving innovation in service delivery and changing the way Service Canada works
- ✚ Many uses of AI will open up new ways of working
- ✚ Transition staff to new skill sets to support AI
- ✚ Increased productivity in the long term
- ✚ Front loaded cost investment in AI

### Web content optimization

A bot is the most disabled person who will visit services online to Canadians, if Service Canada can make it work for the bot it will work for Canadians. A convergence of search engine optimization (SEO) activities is indicated, since AI and humans alike benefit from content optimization in the same format and style.

For example, the Chatbot can be more closely integrated with the Canada.ca website – when providing answers that include a web link, the URL of the browser can be controlled by the machine to take the user to the relevant page on Canada.ca.

- ✚ Content optimization activities on Canada.ca picks up some content being created for Chatbot
- ✚ Siri, Alexa, Google Assistant would all work better with content on Canada.ca that's designed for a chat environment

### Operationalize and Sustain

So far, the Chatbot has been easy to work with however the content needs a lot of time. With additional services being added, focus should be on the content. The main thing is to create the optimized content as a foundational element of the chat space.

- ✚ Leverage the ICM team to populate additional services content in the chat environment
- ✚ Leverage web account managers to influence the Canada.ca pages with the enhanced content
- ✚ Additional work for existing staff needs to be managed in time during service additions
- ✚ Technology license and support fees are required for this chat service
- ✚ Additional work on accessible code is required to operationalize this service
- ✚ Chat search engine could also be used as a training aid for call centre staff



s.20(1)(b)

s.21(1)(a)

s.21(1)(d)

## Lessons Learned

Business need for structured content that can be read aloud

All service content should be formatted to be scanned, and styled to be read out-loud in conversational tone

- ✚ Two line snippets of information that can be used as answers in a conversation
- ✚ Parent-child relationships for all content
- ✚ Reuse structured content for chat in other web spaces

### Publishing Approvals

Service Canada oversight of web pages related to service delivery content should be a requirement going forward. While Service Canada had an easy to use technical solution in place to control the content remotely, coordinating publishing approval across multiple web teams was challenging to manage.

- ✚ Go live is affected by third parties, approvals, etc.
- ✚ Negative testing was conducted, affecting client metrics

### Promotion

Soft launch approach for the initial testing uncovered no show-stoppers. Next time out, the chat should be posted prominently on the page and promoted with keyword links on other pages. Access is required.

- ✚ Service search pages
- ✚ Service contact pages

## Recommendations

### Expand the Build in Canada testing – option 1

The BCIP is finished, however a possible further testing option may be available in approximately six months: The vendor is proposing to apply for BCIP to test new features on its service. If the vendor is successful in its proposal to test new features, ESDC could be matched as a test department again.

### Operationalize the test – option 2

Additional sales vehicle is available through BCIP for ESDC to purchase the service: An annual SLA and monthly fee schedule is available for ongoing technology support.

Further testing is needed to confirm internal resource requirements, these are *provisional numbers*:

Further accessibility code work needed for use in an operational setting, specification is available.

### Conclude the test without further action – option 3

Report on the test and take no further action.



Employment and  
Social Development Canada

Emploi et  
Développement social Canada

Service  
Canada

2017 NHQ 018701

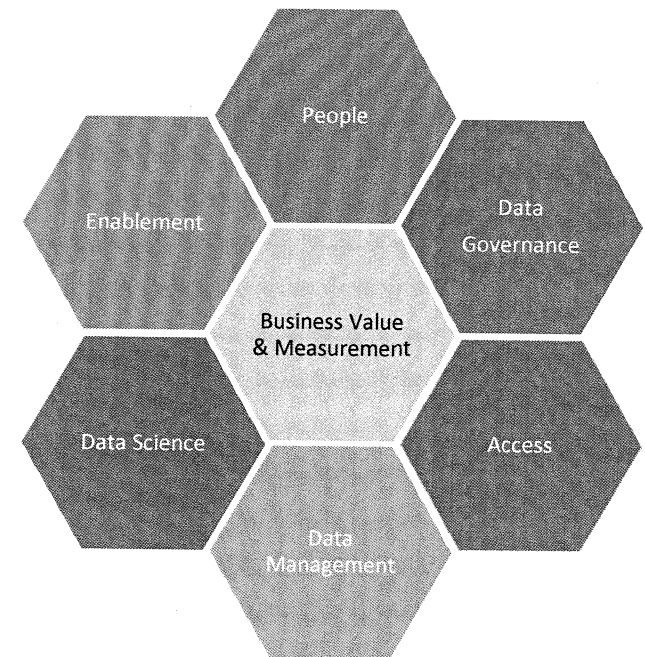
# Data Science

Chief Data Office, SSPB

# Data Science

## Purpose

- Data science is one of the streams of the Data Strategy that aims to provide people with rapid, secure and authorized access to quality data in a way that respects personal privacy and delivers value by giving them the skills, tools, and processes needed to maximize the impact of our enterprise data asset.
- Data science is about using analytic techniques, such as machine learning\*, sentiment analysis and natural language processing, on data to solve business problems.
  - It mines large amounts of data at a granular level to identify complex behaviors, patterns and trends that uncover hidden insights that enable organizations to make smarter decisions.
  - The CDO has led a number of data science pilot projects that are described in slides 10 to 20.



Data Strategy Elements

# Data Science

## Why does ESDC need Data Science?



- ESDC is a heavily manual organization
  - Current methods to process work for clients or generate reports are slow, inefficient, inconsistent, and prone to error
- ESDC has a huge amount of unstructured data (text, scans, audio)
  - A majority of the information held by ESDC is not used because it is hard to access and process using traditional analytical tools
- ESDC's service delivery is too often reactive and not proactive
  - The department is continuously playing catch-up on existing workload inventories. There are millions of items in the backlog and low value items often crowd the critical work. This limits the resources we can put towards more proactive service delivery strategies

By putting data science in place throughout the organization, the CDO Office is enabling ESDC analysts to deliver better results more efficiently, add greater value in their work and provide better service to Canadians.



# Data Science

## How is the CDO Office putting Data Science into place?

- Leverage analytics to maximize the value of data
  - Work with partners on pilot projects to deliver immediate value, while demonstrating future potential
- Develop governance for analytics
  - Provide guidance to the organization on the skills, tools, techniques and processes needed for analytics.
- Democratize analytics
  - Educate on the role and potential of analytics so that groups across the organization can do it independently.
- Provide Expert Advice
  - Support branches in negotiations with external vendors to get the best possible contracts by making sure the department retains all IP from developed models and receives maximum value for its investment.
- Partner with the Innovation, Information and Technology Branch (IITB) to upscale proven data science pilots.
  - Transition pilots into business solutions that integrate Data Science into how we work.

# Where are we right now ?

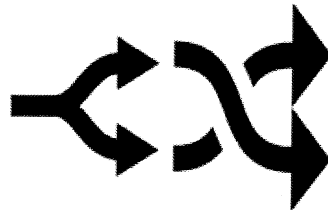


## « Replicate human actions »

Used for rules based, simple to complex process  
Faster handling time  
Higher volumes  
Reduced errors and costs

### ESDC Current Production

- Smart Recommender System
- Workflow triage
- Supported Information Retrieval

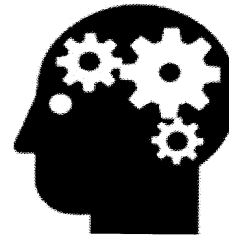


## « Augment human judgement »

- Used for judgement based processing  
- Machine learning capability  
- Interpret human behavior

### ESDC's Pilots

- SCARLET
- T4 Investigating
- Labour question answering machine
- Homelessness and Poverty survey analysis
- Involuntary Separation

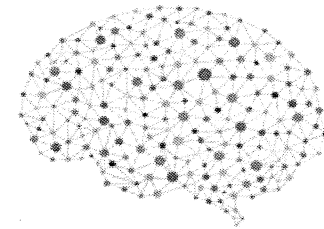


## « Augment human intelligence »

Used for predictive  
« Decisioning »  
Dynamically self-adaptable and managing

### ESDC's Potential

- Client Decision Prediction
- Question answering (Chatbot)
- Speech to text and voice recognition
- Document Drafting



## « Mimics human intelligence »

Acquire human-like thought processing and capabilities

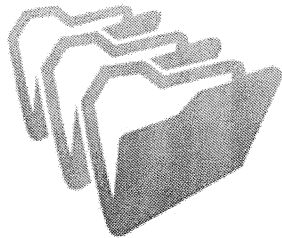
### Where the current big players are:

- |  |  |
|--|--|
| <ul style="list-style-type: none"> <li>• Netflix and Amazon recommendation tools</li> <li>• Tesla's autopilot</li> </ul> | <ul style="list-style-type: none"> <li>• SpaceX rocket landing algorithms</li> <li>• AlphaGo Zero</li> <li>• IBM Watson; Apple devices; Google A.I.</li> </ul> |
|--|--|

# Data Science

## What is the CDO Office using Data Science to do?

### Categorization



Dividing the data into groups, categories, or topics that are related

- Triaging work
- Segmenting clients
- Identifying topics or themes

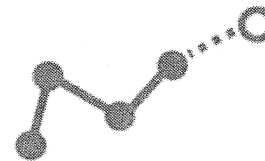
### Key Information Extraction



Isolating important information and extracting the useful part

- Chatbots\*
- Autofill forms or templates
- Descriptive information

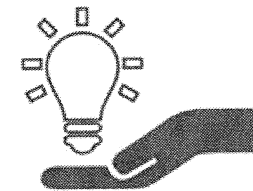
### Prediction



Using historical relationships to predict future outcomes

- Advise agents on decision to be taken
- Predict future workload
- Predict potential benefits for clients

### Technical Expertise to Advise



Exploring emerging technologies and approaches to inform decisions

- Help ESDC sign better contracts
- Identify key emerging tech and adapt it for ESDC's needs.
- Repurpose and adapt models for quick wins

\* See glossary in annex B

# Data Science

## How do machines learn?

- The idea behind machine learning is to train machines the way someone would train a human – through experience and expert guidance
- Supervised learning involves an analyst manually coding examples (a training set) that lets the machine develop an algorithm to mirror those decisions
  - For example, assign a number of documents to group 1, 2, or 3 that allows the machine to see a pattern that it uses to classify future documents
- Unsupervised learning involves the machine being given a defined task without human assisted examples of the “right” and “wrong” answers
  - For example, divide these documents into 4 groups without specifying the groups, based on whatever the computer sees as relevant
- Reinforcement learning is where the machine is able to interact with its environment to determine whether or not it achieved its desired outcome and then adjusts its decision making to improve
  - For example, when Netflix recommends a movie and then the user stops watching after 5 minutes, the algorithm learns that that wasn't a good recommendation and refines its algorithm





# Data Science

## Key CDO Data Science focus: Artificial Intelligence

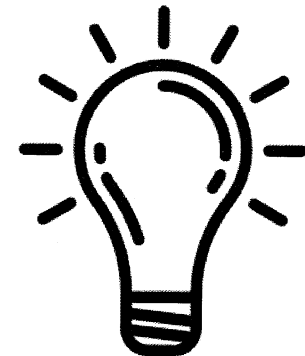
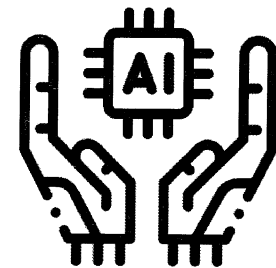
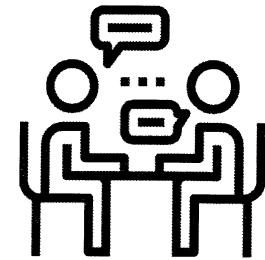
- Leading the Department on A.I.
  - We are developing an **A.I.\* Strategy** for the Department that outlines where we need to go and how we'll get there.
- Building a Departmental A.I. suite
  - In the first project, the CDO is working with the Transformation and Integrated Service Management Branch (TISMB) to implement a workflow triage A.I. that is expected to avoid the department \$500,000 in costs annually.
- Establish A.I. Thought Leadership
  - Support partner branches in negotiations with external vendors, make sure ESDC retains all IP and can properly value the services provided.
  - Working with partner branches to see how A.I. can be beneficial to their work.

\* See glossary in annex B

# Data Science

## Data Science Projects

- The CDO is leading numerous pilot projects with multiple branches in the department.
- The following slides describe a few examples of solutions provided by the CDO that are:
  - Enabling proactive decisions
  - Automating manual processes
  - Leveraging unstructured data (text, image, audio)
  - Preparing for future technologies



# Automating Manual Processes

## Business Context

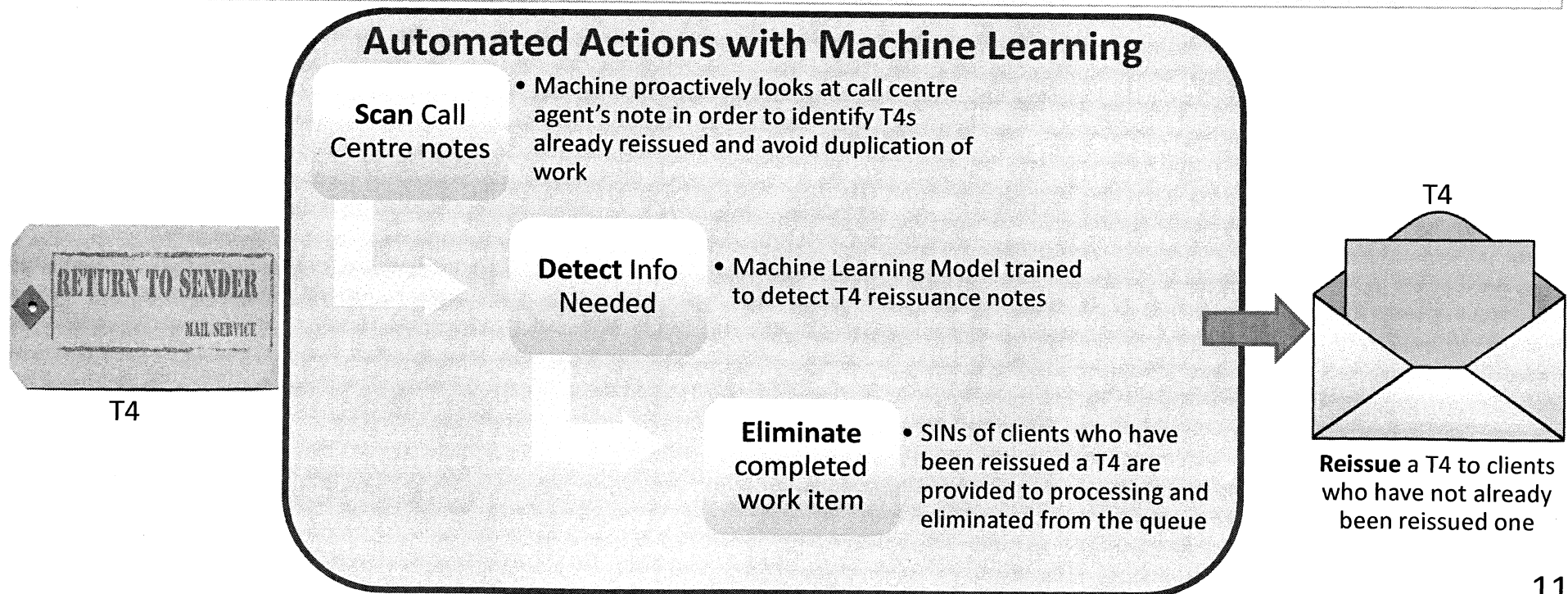
Each year Canada Pension Plan (CPP) and Old Age Security (OAS) recipients receive a T4 form containing the benefits received that they will need to report on their tax return. If they are recipients of the CPP the slip is a T4A(P) while it's a T4A (OAS) for OAS.

## Problem

Each year many T4s are returned to the Service Canada processing network due to address changes or other reasons. A large number of clients previously followed-up with Service Canada to request a duplicate T4, which is immediately reissued. Processing still performs resource intensive investigation of each returned item individually.

## Value

- Low-value work eliminated: 50,000 work items, or 2.5 FTEs, saved annually, while development only used 0.1 FTE
- Work inventory is reduced, with faster processing for remaining clients
- Once built, the models can be repurposed to answer new questions



# Leveraging unstructured data (text, images, audio)

## Business Context

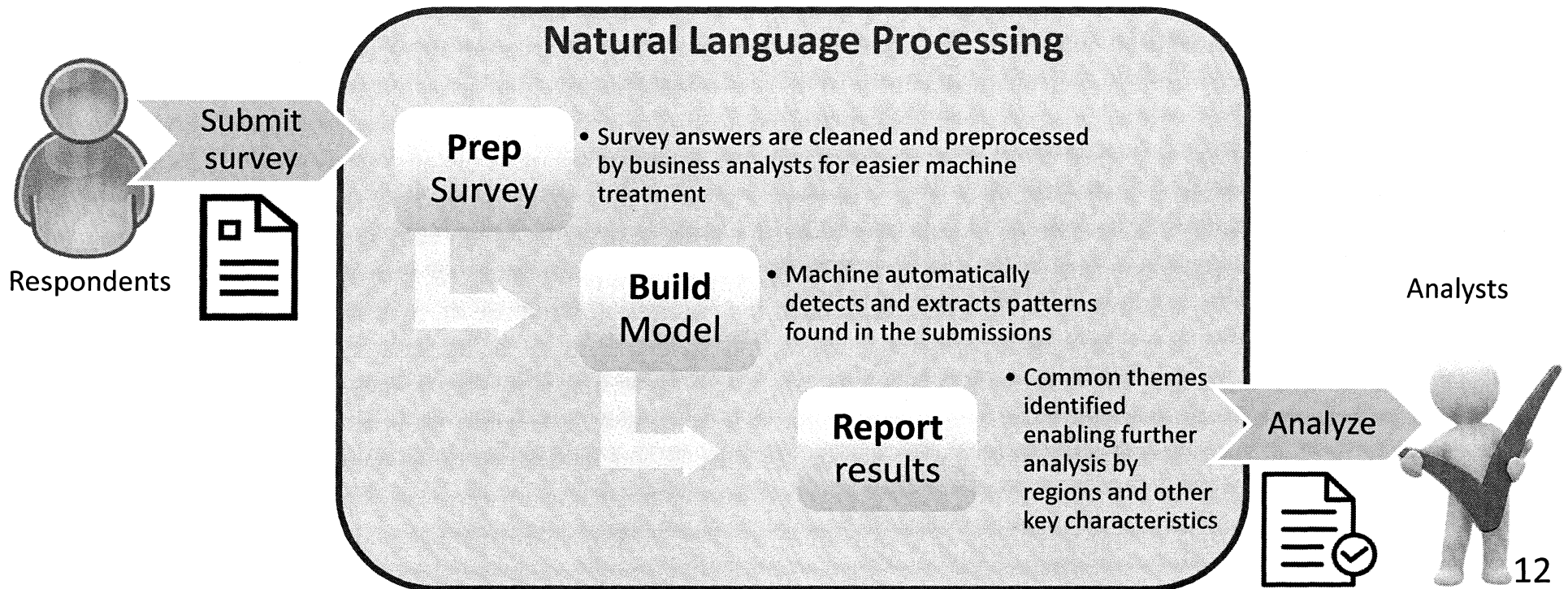
ESDC runs numerous consultations that contain at least one open ended section where participants can express themselves on a specific topic. In 2017, ESDC ran a consultation around the theme of poverty reduction. As part of this consultation, the department received over 800 submissions in free text (emails, online forum and story submissions).

## Problem

To get insight into the Poverty Reduction Strategy, analysts would read through different types of submissions (emails, online forums and surveys) from over 800 respondents and try to extract key information manually. This method is resource intensive and inconsistent.

## Value

- Fast, unbiased and reproducible insights (took one student 3 weeks to build a model of themes)
- Scalable to very large data sets
- Analysts can delve into explaining results rather than identifying trends saving about 1/6 FTE that would have been allocated to reading the submissions.





# Preparing for future technologies

## Business Context

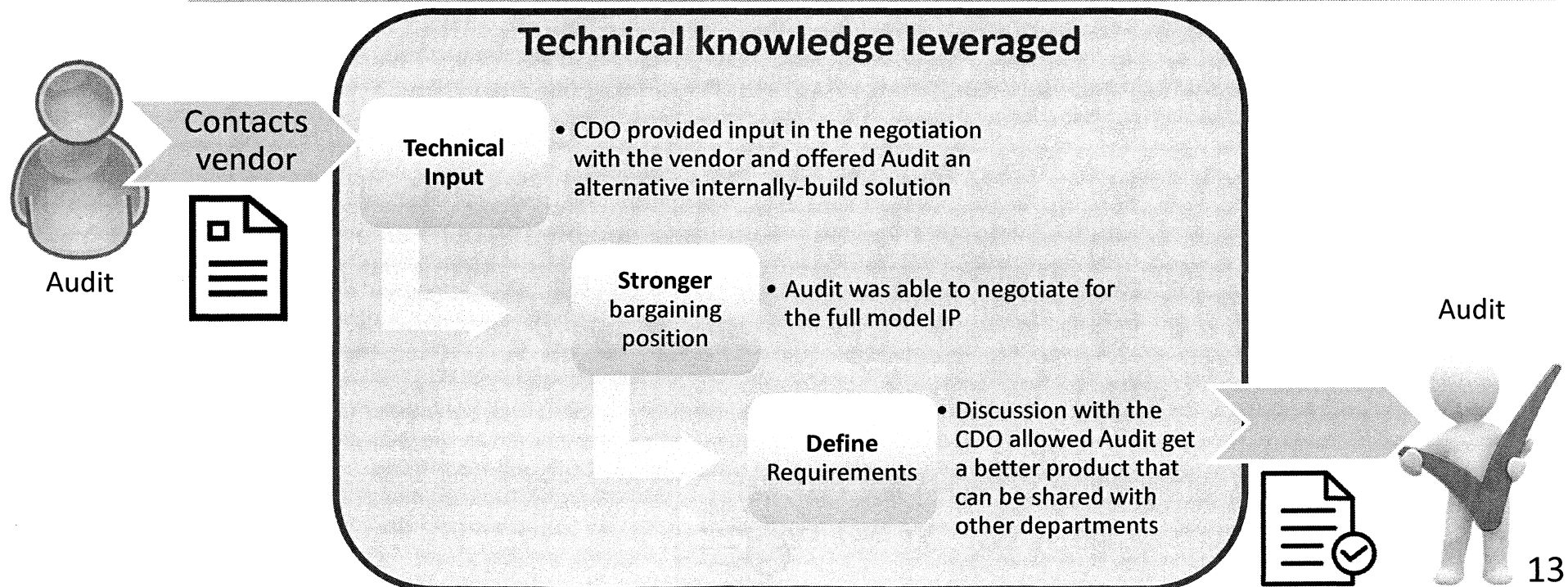
External Artificial Intelligence (AI) tools are increasingly being considered by various branches in their attempt to leverage their data assets. However, the knowledge of the different techniques in the department is sparse so leveraging external vendors is a common option explored.

## Problem

Audit contacted a private vendor to acquire a machine learning categorization tool. The private vendor would have required the department to pay development costs up front, acquire a subscription to use the model and ESDC would own no intellectual property (IP).

## Value

- IP rests with ESDC allowing the model to be repurposed, reused, and shared with other Departments
- Audit contracted for a better product at the same costs
- External expertise is leveraged to hit tight timelines



# Data Science

## Next Steps

- Put models into **production** to deliver value
  - Working closely with the Transformation and Integrated Service Management Branch operations
- Support **Service Transformation Plan** activities
  - Regular participation in the acceleration hub and advising on specific initiatives
- Deliver the first stage of a draft **Analytics Program** in Fall 2017
  - Will be used to engage stakeholders across the organization
- Finalize the **Artificial Intelligence Strategy**
- Work with partners across the organization to discover where data science can help them and **build organizational capacity**
  - Leading data and research streams of 2017-18 post-secondary recruitment to identify strong technical candidates for all of ESDC
- Work with the Innovation, Information and Technology Branch to make **AI sustainable**
  - Developing long-term solution for the models, how they are stored, accessed and maintained

## ANNEX A: Pilot Project Details

Stakeholders	Projects
<b>Human Resources Service Branch</b> <i>(Project Completed)</i>	<p><b>Business Context:</b>            In order to improve employee retention and identify shortcomings in employer-employee relations, HRSB runs an Employee Exit Survey to all employees leaving the department. Some of the fields in this survey are free text.</p> <p><b>Current situation:</b>            Extraction of key information would need to be done manually, which is inconsistent and resources intensive.</p> <p><b>Solution Provided:</b>            Automated the analysis of the employee exit survey using topic modelling. Common themes that may have otherwise fallen through the cracks were identified (i.e. the link between mental health and housing). Using natural language processing is faster, unbiased, scalable and reusable.</p>
<b>Internal Audit Service Branch (IASB)</b> <i>(Project Completed)</i>	<p><b>Business Context:</b>            External Artificial Intelligence (AI) tools are increasingly being considered by various branches in their attempt to leverage their data assets. However, the knowledge of the different techniques in the department is sparse so leveraging external vendors is a common option explored.</p> <p><b>Current Situation:</b>            Audit contacted a private vendor to acquire a machine learning categorization tool. The private vendor would have required the department to pay development costs up front, acquire a subscription to use the model and ESDC would own no intellectual property (IP).</p> <p><b>Solution Provided:</b>            CDO provided input in the negotiation with the vendor and offered Audit an alternative internally-build solution. Audit was able to negotiate for the full model IP with no ongoing subscriptions fees. The discussion between the CDO and Audit allowed Audit to better define their requirements.</p>

Stakeholders	Projects
<p><b>Income Security and Social Development Branch (ISSD) (CPPD)</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b>            Canada Pension Plan (CPP) provides disability benefits to people who have made enough contributions to the CPP and who are disabled and cannot work at any job on a regular basis. Medical adjudicators (MA) determine, after reviewing your application, your eligibility.</p> <p><b>Current Situation:</b>            CPPD applications have to wait for a medical adjudicator's (MA) review to determine if they require additional documentation. Often this review takes place after several months making it hard to meet the service standard of 120 days.</p> <p><b>Solution Provided:</b>            Predicting which client files require additional documentation at receipt to reduce delays and support better and more consistent decisions by CPP-D agents to eliminate costly appeals of decisions.</p>
<p><b>Income Security and Social Development Branch (ISSD) (Homelessness Strategy)</b></p> <p><i>(Project Completed)</i></p>	<p><b>Business Context:</b>            ESDC runs, year after year, numerous consultations which almost always contain at least one open ended section where participants can express themselves on a specific topic. In 2017, ESDC ran a consultation around the theme Homelessness. As part of this consultation, the department received hundreds of submissions in free text.</p> <p><b>Current Situation:</b>            To get insight in the Homelessness Strategy, analysts would read through all submissions and try to extract key information manually. This method is resource intensive, inconsistent and not reproducible.</p> <p><b>Solution Provided:</b>            Through topic modelling, the CDO provided with a fast, scalable and efficient way to extract key information.</p>



Stakeholders	Projects
<p><b>Labour Program – Information Retrieval System</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b>            About 1 million employees in Canada work in federally regulated occupations. The Labour Program promotes cooperation and fairness and provides expert advice and assistance on labour relations matters to workplaces within the federal jurisdiction. To that intent, the Labour Program has developed numerous services, measures and initiatives to assist employers and employees in creating and maintaining a workplace that is conducive to good industrial relations.</p> <p><b>Current Situation:</b>            The labour program, upon requests on collective agreements related to federal jurisdiction occupations, need to extract information manually from a sample of collective agreements because of the large number of agreements.</p> <p><b>Solution Provided:</b>            Creating a question and answer tool to categorize and facilitate responses to queries of collective agreements. Model allows for a faster, more efficient and reproducible way of querying collective agreements. The model required 0.2 FTE to create and can be reused over and over at almost no costs.</p>
<p><b>Legal Services (Justice Canada)</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b>            External A.I. tools are more and more considered by different branches in their attempt to leverage their data assets. Legal is facing an ever growing volume of documentation and need to keep up with the private sector. However, the knowledge of the different techniques in the department is sparse.</p> <p><b>Current Situation:</b>            Legal is interested in acquiring an A.I. solution to keep up with the ever increasing amount of documentation. Legal lacks the proper knowledge to define their requirements in terms of A.I. to external vendors.</p> <p><b>Solution Proposed:</b>            Providing technical expertise to advise on Artificial Intelligence projects with external contractors to facilitate the acquisition of the right solution.</p>

Stakeholders	Projects
<p><b>Strategic and Service Policy Branch (SSPB) – Poverty Reduction Strategy</b></p> <p><i>(Project Completed)</i></p>	<p><b>Business Context:</b> ESDC runs, year after year, numerous consultations which almost always contain at least one open ended section where participants can express themselves on a specific topic. In 2017, ESDC ran a consultation around the theme Poverty Reduction. As part of this consultation, the department received hundreds of submissions in free text.</p> <p><b>Current Situation:</b> To get insight in the Poverty Reduction Strategy, analysts would read through all submissions and try to extract key information manually. This method is resource intensive, inconsistent and not reproducible.</p> <p><b>Solution Provided:</b> Automated the extraction of key themes from hundreds of online submissions related to the Poverty Reduction Strategy.</p>
<p><b>Service Transformation Plan (STP)</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b> ESDC is in the midst of transforming the way we deliver services to Canadians. Among the transformation envisioned are modern authentication methods and automated dialoguing solutions.</p> <p><b>Solution Envisioned:</b> Exploring speech recognition technology to identify how it can be used for both client identification as well as analytics.</p>
<p><b>Transformation and Integrated Service Management Branch (TISMB) – OAS related mail</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b> ESDC is in the midst of transforming the way we deliver services to Canadians. Among the transformation envisioned are modern authentication methods and automated dialoguing solutions.</p> <p><b>Current Situation:</b> A significant amount of emails relating to OAS are hitting the system. Many of those are related to work that cannot be actioned or duplicate work that is currently underway or has been addressed. The current email triaging requires many hours of manual effort, which could otherwise be used in more important areas of processing.</p> <p><b>Solution Provided:</b> Through a number of techniques, the CDO streamlined, closed and aggregated email workflow by triaging emails as they arrived.</p>

Stakeholders	Projects
<p><b>Transformation and Integrated Service Management Branch – OAS overpayment</b></p> <p><i>(Project Completed)</i></p>	<p><b>Business Context:</b> Occasionally under/overpayments are issued to Canadians and Service Canada is notified by the clients. The errors are noted in a system. In order to ensure that every Canadians gets what he/she is entitled too, intensive manual efforts are needed.</p> <p><b>Current Situation:</b> In a number of cases, overpayments and underpayments are issued by ESDC to clients. Information Technology Renewal Delivery System (ITRDS) notes contains information on those over/underpayments. The Transformation and Integrated Service Management Branch employees are manually going through files and extract the information from the ITRDS Notes. This work is time consuming, resource intensive and inconsistent.</p> <p><b>Solution Provided:</b> The CDO used predictive models to flag ITRDS notes reporting an over/underpayment amount over a certain threshold. The model demonstrated accurately that it could extract specific information. The model build by the CDO using only 1/3 FTE is efficient, consistent, is replicable and frees up valuable analysts time. Also, the model developed here can be repurposed to do similar task internally at ESDC or in other departments facing similar issues.</p>
<p><b>Transformation and Integrated Service Management Branch – T4 Processing</b></p> <p><i>(Project Completed)</i></p>	<p><b>Business Context:</b> Each year Canada Pension Plan (CPP) and Old Age Security (OAS) recipients receive an information slip containing the information they will need to report on their tax return. If they are recipients of the CPP the slip is a T4A(P) while it's a T4A (OAS) for OAS.</p> <p><b>Current Situation:</b> Service Canada processing network receives numerous returned T4's due to changes in client's address or for other reasons. A significant number of clients follow-up with Service Canada to request a duplicate tax slip. Processing is not made aware which in return create intensive manual investigation.</p> <p><b>Solution Provided:</b> The model identifies when a client's T4 as being reissued, provided the SIN associated with that T4 to processing and eliminates it from the queue. 50,000 work items, or 2.5 FTEs, saved annually, while development only used 0.1 FTE. Work inventory is reduced, with faster processing for remaining clients. Once built, the models can be repurposed to answer new questions.</p>



Stakeholders	Projects
<p><b>Transformation and Integrated Service Management Branch – EI Application</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b>            ESDC is still relying heavily on paper application for the numerous programs in its portfolio. Among them, the Employment Insurance (EI) program requires Canadians to fill a paper application.</p> <p><b>Current Situation:</b>            EI application are manually scanned and tags are manually assigned to categorized them.</p> <p><b>Solution Proposed:</b>            Automating the classification of EI applications allowing the reallocation of valuable employee time.</p>
<p><b>Transformation and Integrated Service Management Branch – Involuntary Separation – GIS</b></p> <p><i>(Project Completed)</i></p>	<p><b>Business Context:</b>            The GIS provides a monthly non-taxable benefit to Old Age Security (OAS) pension recipients who have low income and are living in Canada. In January 2017, guidelines for Guaranteed Income Supplement (GIS) were changed. The GIS guidelines, in the case of a partner admitted to long-term care facility would no longer allow those affected to be assessed as 'single'. This situation leaves one person to pay both household costs and long-term care facility expenses with less GIS support. The Minister of Families, Children and Social Development announced that the guidelines were to be changed to what they were prior to January 2017.</p> <p><b>Current Situation:</b>            After a policy change, a number of applicants were denied GIS. The change in policy was reversed and analysts would need to go through all . Information Technology Renewal Delivery System (ITRDS) notes to identify the clients that were denied.</p> <p><b>Solution Provided:</b>            The CDO built a model to identify those automatically, saving numerous FTEs that can be allocated to other tasks.</p>
<p><b>Innovation, Information and Technology Branch (IITB), Shared Service Canada (SSC) (Data Lake)</b></p> <p><i>(Project Underway)</i></p>	<p><b>Current Situation:</b>            ESDC analysts and external researchers requiring protected data must go through a lengthy and complicated process to access data.</p> <p><b>Solution Proposed:</b>            Exploring a Protected B Cloud Pilot on a variety of data sources including image and speech to provide easier access and advanced analytical tools to researchers.</p>

Stakeholders	Projects
<p><b>Public Affairs and Stakeholder Relations Branch (PASRB) – Media Monitoring</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b>          Everyday, thousands of articles and news stories on ESDC or its Ministers are produced. Keeping track of what is being said about the department, its programs or its Ministers with manual techniques is becoming more and more difficult as the speed at which news are produced.</p> <p><b>Current Situation:</b>          The Government of Canada (GoC), through Public Safety Canada, purchases a list of articles pertaining to all federal departments. Those lists are then shared with departments for further analysis. ESDC employees currently come in to work at 3 A.M. in order to open and read between 800 and 1000 articles. Of those, about 100 are evaluated and summarized for ESDC's higher management. Manual monitoring of a sample of all news related to the department is resource intensive and time consuming.</p> <p><b>Solution Proposed:</b>          The CDO is working on a tool that will rate not only a sample, but all the articles provided by Public Safety Canada everyday. It will rate articles according to their content, the mention of an ESDC program or if ESDC's Ministers are mentioned. Once the monitoring part of the model is up and running, an automatic summarizer will be implemented.</p>

NOW AND TOMORROW EXCELLENCE IN EVERYTHING WE DO



# Data Science

Chief Data Office, SSPB



Employment and  
Social Development Canada

Emploi et  
Développement social Canada

Canada



# The world of data has changed, and it means change for ESDC

- **Client expectations have changed**

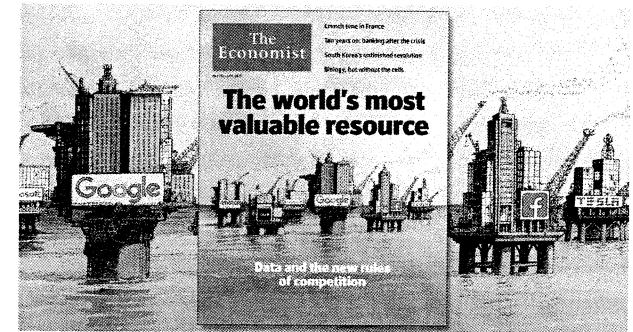
Clients expect private sector level of service, but with a much higher standard for privacy protection. We must integrate our data to support proactive client service within secure and managed environments.

- **ESDC has changed**

Evidence-based decision making and Results & Delivery, the Service Transformation mandate, and a focus on Transparency require access to data. We must know what data we have and how to use it.

- **Technology and Analytical methods have changed**

We have fallen behind in the underlying investments needed to use and extract the value of data. We need people, technology and an analytics program to tie it all together.



**"I believe that government departments and organizations urgently need to turn their attention to this issue. They need to focus on collecting the right data to support their activities, on ensuring that data is well-managed and up-to-date, and on fully using this data not only to inform their core business, but also to support reporting and continuous improvement."**

**2016 Spring Reports of the Auditor General of Canada, Opening Statement, May 3, 2016**



## ESDC's Data Strategy builds the foundations for improved client service and policy through six work streams

Getting data to everyone who needs it quickly and securely while protecting privacy

Access

Data Science

Building a program to develop analytical capacity to uncover new insights from ESDC's data

### Foundations for leveraging data and analytics

Data  
Governance

Knowing what data we have; what we need; where it is; if we can trust it; who makes decisions; and what rules should be applied

Data  
Management

Ensuring infrastructure is in place to securely store and provide users with access to data and with the tools they need to analyse it

People

Recruiting, retaining and engaging with the people with the skills and experience we need

Enablement

Empowering people to use data and to innovate





Data  
Science

## What challenges are addressed by Data Science?

- ESDC is still a heavily manual organization
  - We have made progress, but more changes are needed to make current methods of processing work for clients, generating reports, or finding answers to questions faster, more efficient, more consistent, and less prone to error
- ESDC has a huge amount of unstructured data
  - A majority of the information held by ESDC is not used because it is hard to access and process using traditional analytical tools
- ESDC is reactive not proactive
  - Currently struggling to meet existing and historical demands which prevents us from focusing on what we should be doing to improve

To meet our need for Data Science, ESDC's Chief Data Office is building a program to develop analytical capacity for using methods such as machine learning, Artificial Intelligence, and others to uncover new insights from ESDC's data.

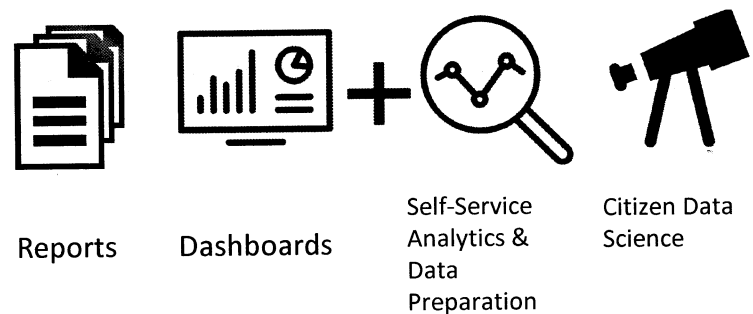


## How is ESDC putting Data Science into place?

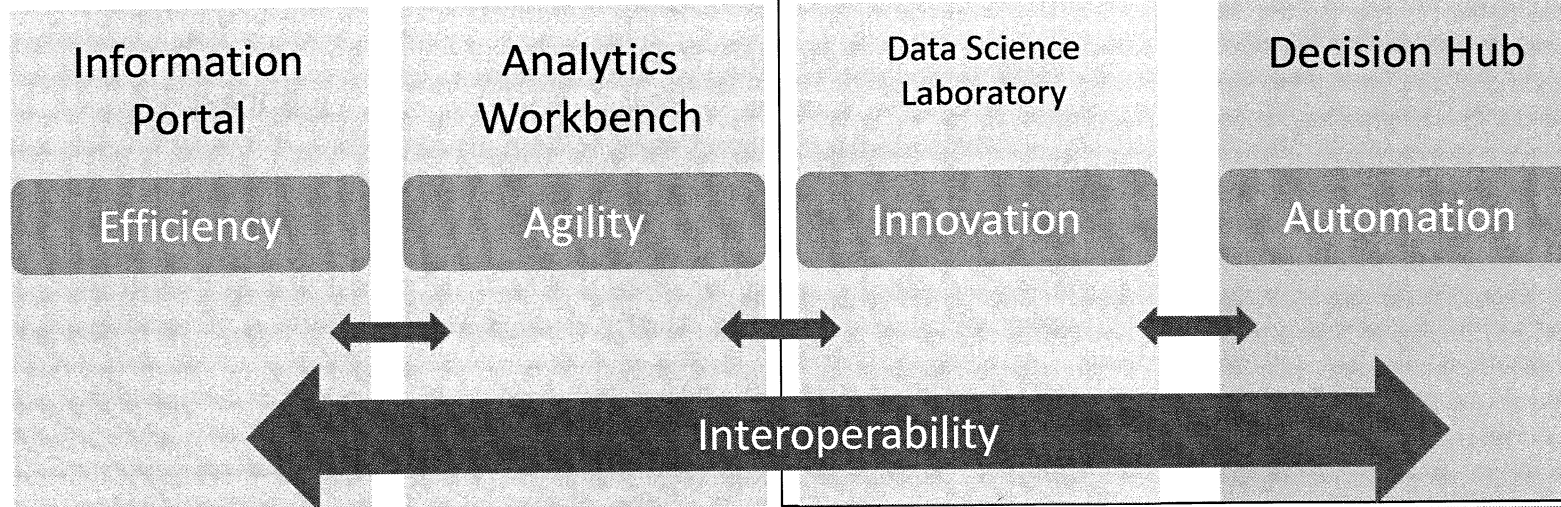
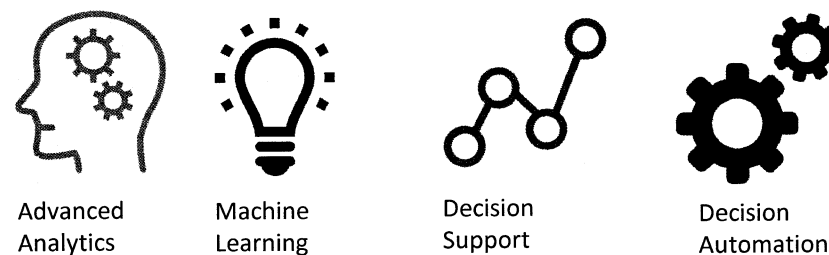
- Leverage analytics to maximize the value of data
  - Work with internal partners on pilot projects to deliver immediate value, while demonstrating future potential
- Develop governance for analytics
  - Provide guidance to the organization on the skills, tools, techniques and processes needed for analytics.
- Democratize analytics
  - Educate on the role and potential of analytics so that groups across the organization can do it independently.
- Provide Expert Advice
  - Support branches in negotiations with external vendors to get the best possible contracts by making sure the department retains all IP from developed models and receives maximum value for its investment.
- A partnership between the Chief Data Office and the Innovation, Information and Technology Branch (IITB) to upscale proven data science pilots.
  - Transition pilots into business solutions that integrate Data Science into how we work.



# Spectrum of Analytics



## Algorithmic Government



## Key Data Science focus within ESDC's Chief Data Office: Artificial Intelligence

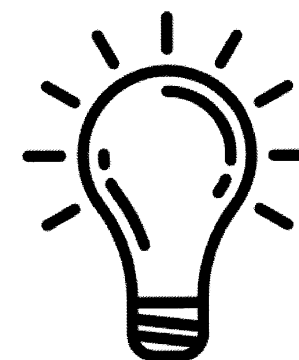
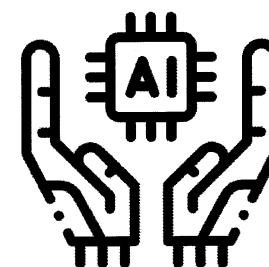
- Leading the Department on A.I.
  - We are developing an **A.I.\* Strategy** for the Department that outlines where we need to go and how we'll get there.
- Building a Departmental A.I. suite
  - In the first project, the CDO is working with the Transformation and Integrated Service Management Branch (TISMB) to implement a workflow triage A.I. that is expected to save the department \$500,000 annually.
- Establish A.I. Thought Leadership
  - Support partner branches in negotiations with external vendors, make sure ESDC retains all IP and can properly value the services provided.
  - Working with partner branches to see how A.I. can be beneficial to their work.

\* See glossary in annex



# Data Science Projects

- A number of pilot projects are currently underway with multiple branches in the department.
- The following slides describe solutions provided by the CDO that are:
  - Enabling proactive decisions
  - Automating manual processes
  - Leveraging unstructured data (text, image, audio)
  - Preparing for future technologies
- Detailed summaries of all the pilot projects completed and underway is also provided in an annex





11

## Labour Example

### Business Context

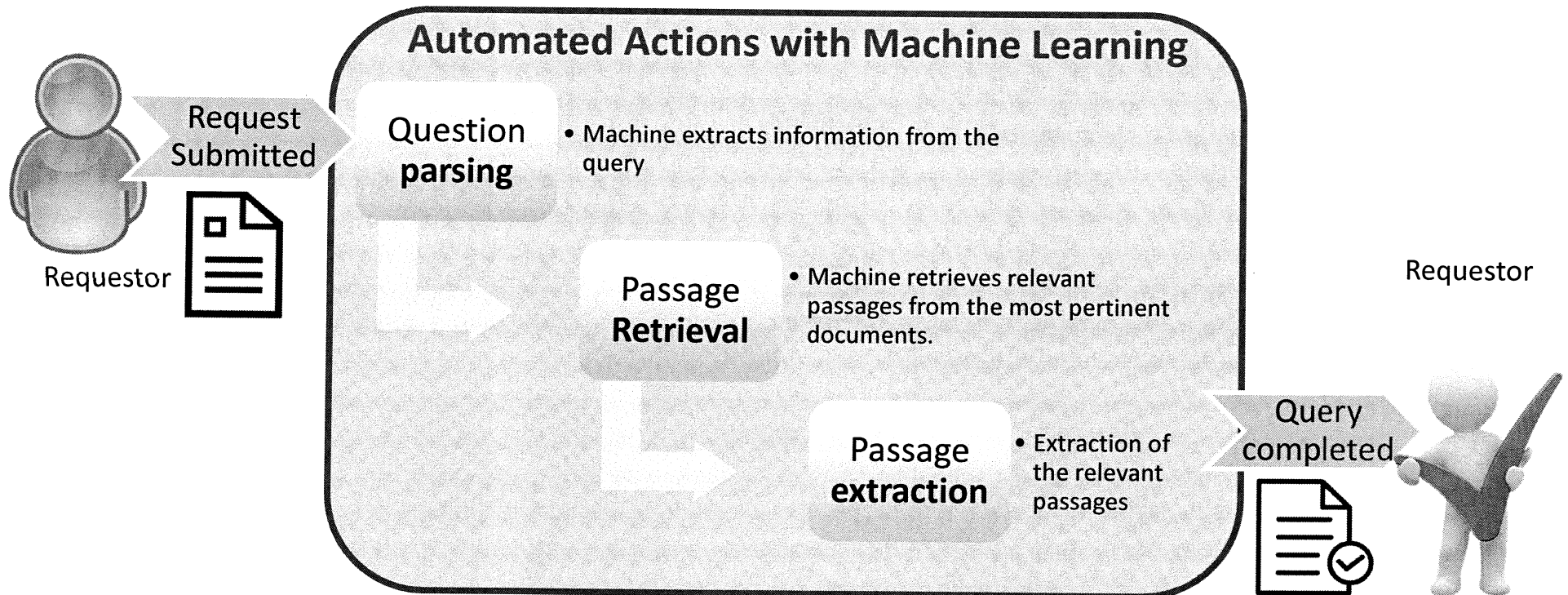
About 1 million employees in Canada work in federally regulated occupations. The Labour Program promotes cooperation and fairness and provides expert advice and assistance on labour relations matters to workplaces within the federal jurisdiction. To that intent, the Labour Program has developed numerous services, measures and initiatives to assist employers and employees in creating and maintaining a workplace that is conducive to good industrial relations.

### Problem

The labour program, upon requests on collective agreements related to federal jurisdiction occupations, need to extract information manually from a sample of collective agreements because of the large number of agreements (+50,000 collective agreements).

### Value

- Faster querying system
- Research is done on all the collective agreements, not a sample
- Model can be reuse on other similar question-answer problems



# Enabling proactive decisions

12

Business  
Context

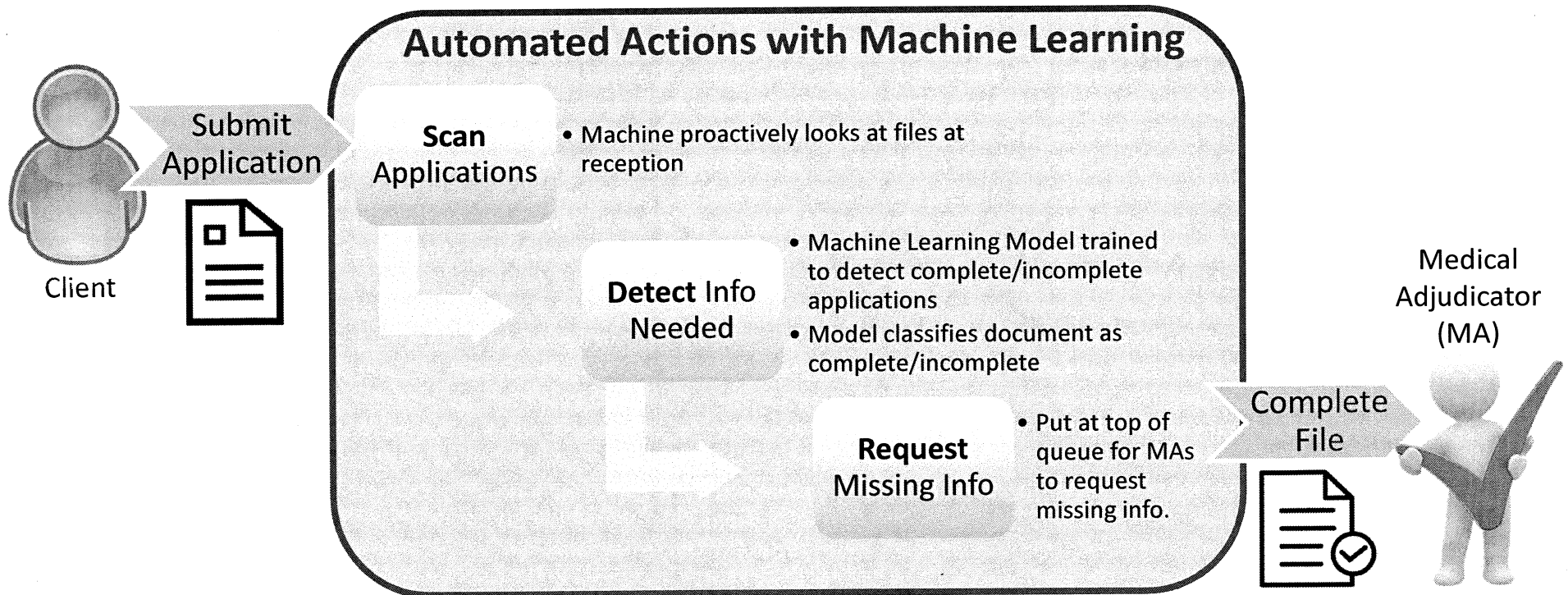
The Canada Pension Plan (CPP) provides disability benefits to people who are disabled and cannot work at any job on a regular basis through the CPPD program. Medical adjudicators (MA) determine eligibility after reviewing detailed applications involving significant medical documentation

Problem

Applicants have to wait for a MA review to determine if they require additional documentation. Often the MA will not look at an application for several months, making it difficult to meet the service standard of 120 days when extra information is needed.

Value

- Clients get their benefits faster
- Fewer rejections and appeals due to missing information
- MA's time more efficiently used to evaluate complete files



# Automating Manual Processes

13  
Business  
Context

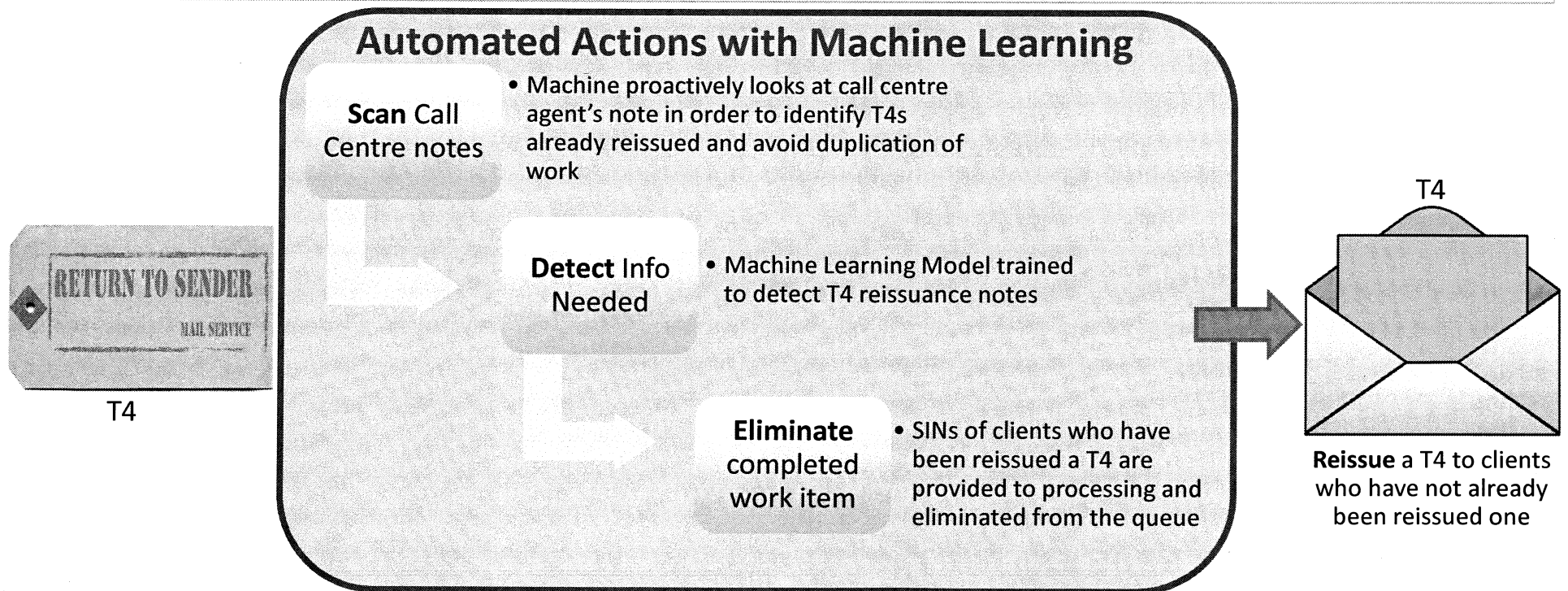
Each year Canada Pension Plan (CPP) and Old Age Security (OAS) recipients receive a T4 form containing the benefits received that they will need to report on their tax return. If they are recipients of the CPP the slip is a T4A(P) while it's a T4A (OAS) for OAS.

Problem

Each year many T4s are returned to the Service Canada processing network due to address changes or other reasons. A large number of clients previously followed-up with Service Canada to request a duplicate T4, which is immediately reissued. Processing still performs resource intensive investigation of each returned item individually.

Value

- Low-value work eliminated: 50,000 work items, or 2.5 FTEs, saved annually, while development only used 0.1 FTE
- Work inventory is reduced, with faster processing for remaining clients
- Once built, the models can be repurposed to answer new questions





14

## Leveraging unstructured data (text, images, audio)

Business  
Context

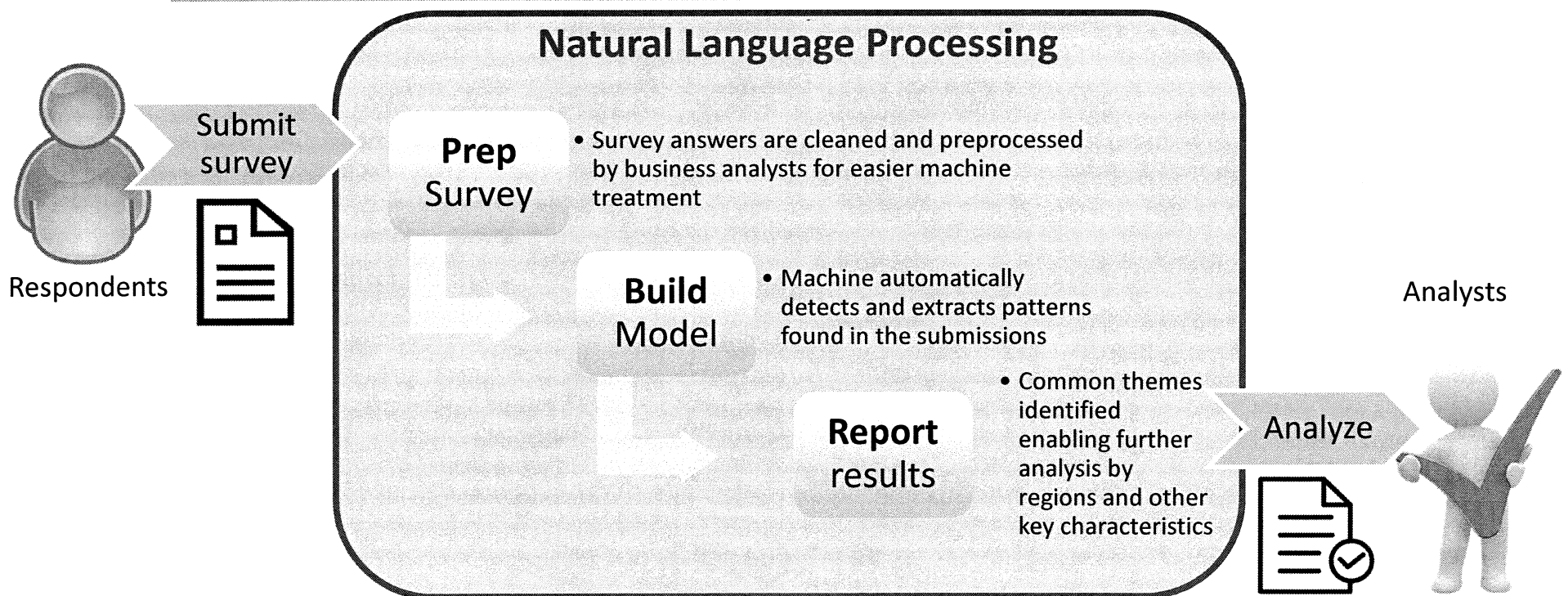
ESDC runs numerous consultations that contain at least one open ended section where participants can express themselves on a specific topic. In 2017, ESDC ran a consultation around the theme of poverty reduction. As part of this consultation, the department received over 800 submissions in free text (emails, online forum and story submissions).

Problem

To get insight into the Poverty Reduction Strategy, analysts would read through different types of submissions (emails, online forums and surveys) from over 800 respondents and try to extract key information manually. This method is resource intensive and inconsistent.

Value

- Fast, unbiased and reproducible insights (took one student 3 weeks to build a model of themes)
- Scalable to very large data sets
- Analysts can delve into explaining results rather than identifying trends saving about 1/6 FTE that would have been allocated to reading the submissions.



15

## Preparing for future technology

### Business Context

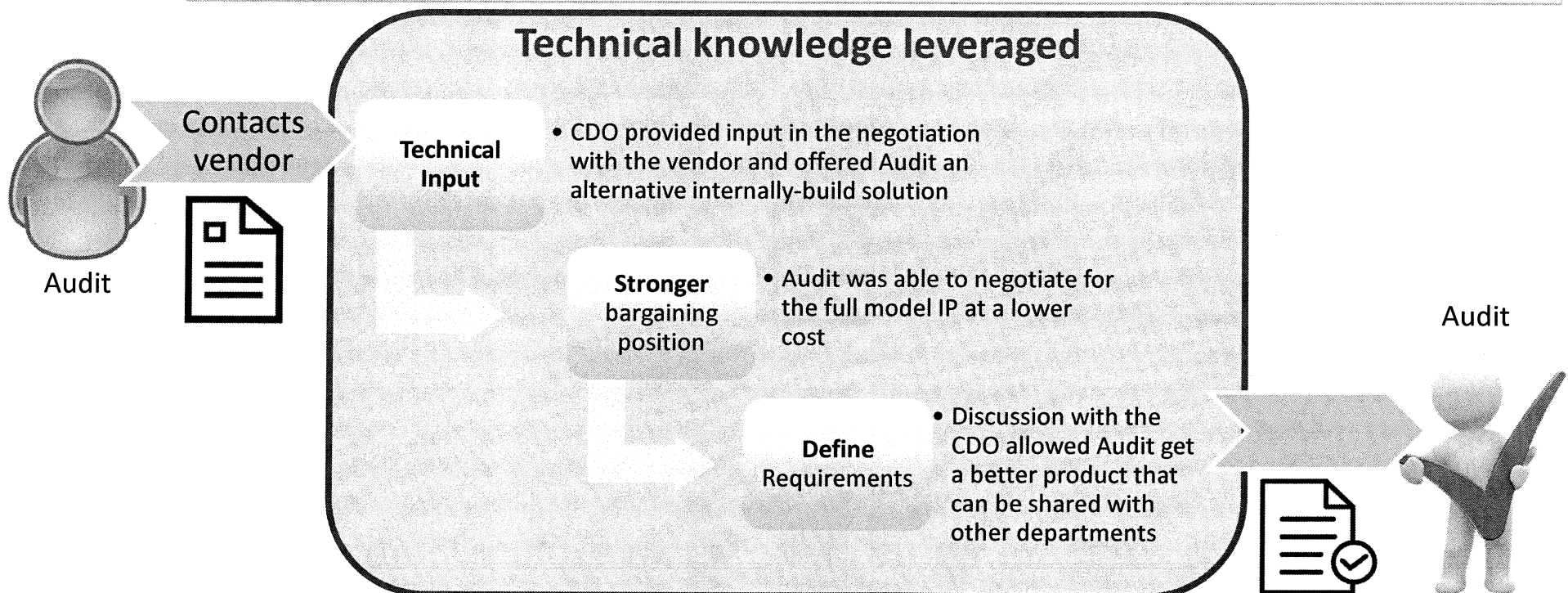
External Artificial Intelligence (AI) tools are increasingly being considered by various branches in their attempt to leverage their data assets. However, the knowledge of the different techniques in the department is sparse so leveraging external vendors is a common option explored.

### Problem

Audit contacted a private vendor to acquire a machine learning categorization tool. The private vendor would have required the department to pay development costs up front, acquire a subscription to use the model and ESDC would own no intellectual property (IP).

### Value

- IP rests with ESDC allowing the model to be repurposed, reused, and shared with other Departments
- Audit contracted for a better product at the same costs
- External expertise is leveraged to hit tight timelines



# Monitoring Impact with Information Extraction

16

Business  
Context

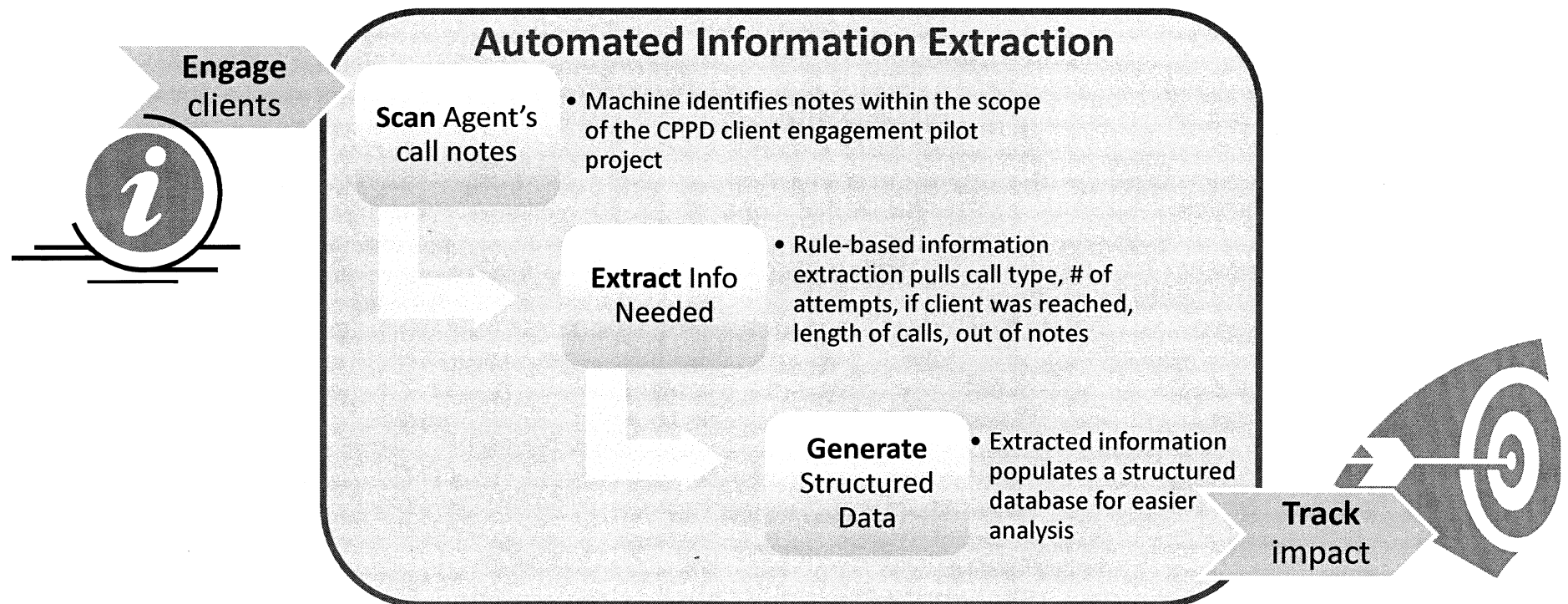
Canada Pension Plan Disability (CPPD) provides benefits to Canadians who are eligible based on medical and non-medical criteria. To reduce costly appeals of denied benefits, a pilot project is engaging clients with 'pre-decision calls' allowing a final chance to provide information to reconsider a denial decision.

Problem

If client engagement causes a significant increase in grants, or reduction in appeals, there are direct implications for budget and operations. The pilot and its efficacy need to be tracked, but the CPPD application process is already complex and labour intensive for the agents.

Value

- Automatic generation of structured data from free text allows agents to record call notes as per usual without extra documentation work
- Rule-based extraction targets specific information needed (e.g. call type, number of calls) without expensive labelling of data by humans



## Next Steps

- Put models into **production** to deliver value
  - Working closely with the Transformation and Integrated Service Management Branch operations
- Support **Service Transformation Plan** activities
  - Regular participation in the acceleration hub and advising on specific initiatives
- Establish **Analytics Program** throughout 2018
  - Built in collaboration with stakeholders across the organization
- Finalize the **Artificial Intelligence Strategy**
- Work with partners across the organization to discover where data science can help them and **build organizational capacity**
  - Leading data and research streams of 2017-18 post-secondary recruitment to identify strong technical candidates for all of ESDC
- Work with the Innovation, Information and Technology Branch to make **AI sustainable**
  - Developing long-term solution for the models, how they are stored, accessed and maintained





## ANNEX A: Glossary

- **Data science** uses a wide-range of analytic techniques on large amounts of granular data to solve business problems.
- **Advanced Analytics** is very similar to data science but focuses on the techniques rather than the overall problem-solving function.
- **Artificial Intelligence** refers to the ability of computers to complete tasks and make decisions that require human-level judgement. Current AI often makes use of machine learning.
- **Chatbot** refers to an interactive digital question and answer tool.
- **Machine Learning** refers to computer algorithms that are able to learn how to solve specific problems through exposure to data and can improve over time as more data is acquired.
- **Natural language processing** refers to computer algorithms that deal with the intake, interpretation, summarization and discourse of natural language (both written and spoken).
- **Sentiment Analysis** refers to the use of algorithms to identify and extract the emotional reaction of the speaker or writer to an event or a document.



## ANNEX B: How do machines learn?

- The idea behind machine learning is to train machines the way someone would train a human – through experience and expert guidance
- Supervised learning involves an analyst manually coding examples (a training set) that lets the machine develop an algorithm to mirror those decisions
  - For example, assign a number of documents to group 1, 2, or 3 that allows the machine to see a pattern that it uses to classify future documents
- Unsupervised learning involves the machine being given a defined task without human assisted examples of the “right” and “wrong” answers
  - For example, divide these documents into 4 groups without specifying the groups, based on whatever the computer sees as relevant
- Reinforcement learning is where the machine is able to interact with its environment to determine whether or not it achieved its desired outcome and then adjusts its decision making to improve
  - For example, when Netflix recommends a movie and then the user stops watching after 5 minutes, the algorithm learns that that wasn't a good recommendation and refines its algorithm



## Annex C: Description of projects by stakeholders (1 of 8)

Stakeholders	Projects
<b>Human Resources Service Branch</b>  <i>(Project Completed)</i>	<p><b>Business Context:</b> In order to improve employee retention and identify shortcomings in employer-employee relations, HRSB runs an Employee Exit Survey to all employees leaving the department. Some of the fields in this survey are free text.</p> <p><b>Current situation:</b> Extraction of key information would need to be done manually, which is inconsistent and resources intensive.</p> <p><b>Solution Provided:</b> Automated the analysis of the employee exit survey using topic modelling. Common themes that may have otherwise fallen through the cracks were identified (i.e. the link between mental health and housing). Using natural language processing is faster, unbiased, scalable and reusable.</p>
<b>Internal Audit Service Branch (IASB)</b>  <i>(Project Completed)</i>	<p><b>Business Context:</b> External Artificial Intelligence (AI) tools are increasingly being considered various branches in their attempt to leverage their data assets. However, the knowledge of the different techniques in the department is sparse so leveraging external vendors is a common option explored.</p> <p><b>Current Situation:</b> Audit contacted a private vendor to acquire a machine learning categorization tool. The private vendor would have required the department to pay development costs up front, acquire a subscription to use the model and ESDC would own no intellectual property (IP).</p> <p><b>Solution Provided:</b> CDO provided input in the negotiation with the vendor and offered Audit an alternative internally-build solution. Audit was able to negotiate for the full model IP with no ongoing subscriptions fees. The discussion between the CDO and Audit allowed Audit to better define their requirements.</p>

## Annex C: Description of projects by stakeholders (2 of 8)

Stakeholders	Projects
<b>Income Security and Social Development Branch (ISSD) (CPPD)</b>  <i>(Project Underway)</i>	<p><b>Business Context:</b> Canada Pension Plan (CPP) provides disability benefits to people who have made enough contributions to the CPP and who are disabled and cannot work at any job on a regular basis. Medical adjudicators (MA) determine, after reviewing your application, your eligibility.</p> <p><b>Current Situation:</b> CPPD applications have to wait for a medical adjudicator's (MA) review to determine if they require additional documentation. Often this review takes place after several months making it hard to meet the service standard of 120 days.</p> <p><b>Solution Provided:</b> Predicting which client files require additional documentation at receipt to reduce delays and support better and more consistent decisions by CPP-D agents to eliminate costly appeals of decisions.</p>
<b>Income Security and Social Development Branch (ISSD) (Homelessness Strategy)</b>  <i>(Project Completed)</i>	<p><b>Business Context:</b> ESDC runs, year after year, numerous consultations which almost always contain at least one open ended section where participants can express themselves on a specific topic. In 2017, ESDC ran a consultation around the theme Homelessness. As part of this consultation, the department received hundreds of submissions in free text.</p> <p><b>Current Situation:</b> To get insight in the Homelessness Strategy, analysts would read through all submissions and try to extract key information manually. This method is resource intensive, inconsistent and not reproducible.</p> <p><b>Solution Provided:</b> Through topic modelling, the CDO provided with a fast, scalable and efficient way to extract key information.</p>



## Annex C: Description of projects by stakeholders (3 of 8)

Stakeholders	Projects
<p><b>Labour Program – Information Retrieval System</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b> About 1 million employees in Canada work in federally regulated occupations. The Labour Program promotes cooperation and fairness and provides expert advice and assistance on labour relations matters to workplaces within the federal jurisdiction. To that intent, the Labour Program has developed numerous services, measures and initiatives to assist employers and employees in creating and maintaining a workplace that is conducive to good industrial relations.</p> <p><b>Current Situation:</b> The labour program, upon requests on collective agreements related to federal jurisdiction occupations, need to extract information manually from a sample of collective agreements because of the large number of agreements.</p> <p><b>Solution Provided:</b> Creating a question and answer tool to categorize and facilitate responses to queries of collective agreements. Model allows for a faster, more efficient and reproducible way of querying collective agreements. The model required 0.2 FTE to create and can be reused over and over at almost no costs.</p>
<p><b>Legal Services (Justice Canada)</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b> External A.I. tools are more and more considered by different branches in their attempt to leverage their data assets. Legal is facing an ever growing volume of documentation and need to keep up with the private sector. However, the knowledge of the different techniques in the department is sparse.</p> <p><b>Current Situation:</b> Legal is interested in acquiring an A.I. solution to keep up with the ever increasing amount of documentation. Legal lacks the proper knowledge to define their requirements in terms of A.I. to external vendors.</p> <p><b>Solution Proposed:</b> Providing technical expertise to advise on Artificial Intelligence projects with external contractors to facilitate the acquisition of the right solution.</p>

## Annex C: Description of projects by stakeholders (4 of 8)

Stakeholders	Projects
<b>Strategic and Service Policy Branch (SSPB) – Poverty Reduction Strategy</b>  <i>(Project Completed)</i>	<p><b>Business Context:</b>  ESDC runs, year after year, numerous consultations which almost always contain at least one open ended section where participants can express themselves on a specific topic. In 2017, ESDC ran a consultation around the theme Poverty Reduction. As part of this consultation, the department received hundreds of submissions in free text.</p> <p><b>Current Situation:</b>  To get insight in the Poverty Reduction Strategy, analysts would read through all submissions and try to extract key information manually. This method is resource intensive, inconsistent and not reproducible.</p> <p><b>Solution Provided:</b>  Automated the extraction of key themes from hundreds of online submissions related to the Poverty Reduction Strategy.</p>
<b>Service Transformation Plan (STP)</b>  <i>(Project Underway)</i>	<p><b>Business Context:</b>  ESDC is in the midst of transforming the way we deliver services to Canadians. Among the transformation envisioned are modern authentication methods and automated dialoguing solutions.</p> <p><b>Solution Envisioned:</b>  Exploring speech recognition technology to identify how it can be used for both client identification as well as analytics.</p>
<b>Transformation and Integrated Service Management Branch (TISMB) – OAS related mail</b>  <i>(Project Underway)</i>	<p><b>Business Context:</b>  ESDC is in the midst of transforming the way we deliver services to Canadians. Among the transformation envisioned are modern authentication methods and automated dialoguing solutions.</p> <p><b>Current Situation:</b>  A significant amount of emails relating to OAS are hitting the system. Many of those are related to work that cannot be actioned or duplicate work that is currently underway or has been addressed. The current email triaging requires many hours of manual effort, which could otherwise be used in more important areas of processing.</p> <p><b>Solution Provided:</b>  Through a number of techniques, the CDO streamlined, closed and aggregated email workflow by triaging emails as they arrived.</p>

## Annex C : Description of projects by stakeholders (5 of 8)

Stakeholders	Projects
<p><b>Transformation and Integrated Service Management Branch – OAS overpayment</b></p> <p><i>(Project Completed)</i></p>	<p><b>Business Context:</b> Occasionally under/overpayments are issued to Canadians and Service Canada is notified by the clients. The errors are noted in a system. In order to ensure that every Canadian gets what he/she is entitled to, intensive manual efforts are needed.</p> <p><b>Current Situation:</b> In a number of cases, overpayments and underpayments are issued by ESDC to clients. Information Technology Renewal Delivery System (ITRDS) Notes contains information on those over/underpayments. The Transformation and Integrated Service Management Branch employees are manually going through files and extract the information from the ITRDS Notes. This work is time consuming, resource intensive and inconsistent.</p> <p><b>Solution Provided:</b> The CDO used predictive models to flag ITRDS Notes reporting an over/underpayment amount over a certain threshold. The model demonstrated accurately that it could extract specific information. The model build by the CDO using only 1/3 FTE is efficient, consistent, is replicable and frees up valuable analysts time. Also, the model developed here can be repurposed to do similar task internally at ESDC or in other departments facing similar issues.</p>
<p><b>Transformation and Integrated Service Management Branch – T4 Processing</b></p> <p><i>(Project Completed)</i></p>	<p><b>Business Context:</b> Each year Canada Pension Plan (CPP) and Old Age Security (OAS) recipients receive an information slip containing the information they will need to report on their tax return. If they are recipients of the CPP the slip is a T4A(P) while it's a T4A (OAS) for OAS.</p> <p><b>Current Situation:</b> Service Canada processing network receives numerous returned T4's due to changes in client's address or for other reasons. A significant number of clients follow-up with Service Canada to request a duplicate tax slip. Processing is not made aware which in return create intensive manual investigation.</p> <p><b>Solution Provided:</b> The model identifies when a client's T4 as being reissued, provided the SIN associated with that T4 to processing and eliminates it from the queue. 50,000 work items, or 2.5 FTEs, saved annually, while development only used 0.1 FTE. Work inventory is reduced, with faster processing for remaining clients. Once built, the models can be repurposed to answer new questions.</p>

## Annex C : Description of projects by stakeholders (6 of 8)

Stakeholders	Projects
<p><b>Transformation and Integrated Service Management Branch – EI Application</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b> ESDC is still relying heavily on paper application for the numerous programs in its portfolio. Among them, the Employment Insurance (EI) program requires Canadians to fill a paper application.</p> <p><b>Current Situation:</b> EI application are manually scanned and tags are manually assigned to categorized them.</p> <p><b>Solution Proposed:</b> Automating the classification of EI applications allowing the reallocation of valuable employee time.</p>
<p><b>Transformation and Integrated Service Management Branch – Involuntary Separation – GIS</b></p> <p><i>(Project Completed)</i></p>	<p><b>Business Context:</b> The GIS provides a monthly non-taxable benefit to Old Age Security (OAS) pension recipients who have low income and are living in Canada. In January 2017, guidelines for Guaranteed Income Supplement (GIS) were changed. The GIS guidelines, in the case of a partner admitted to long-term care facility would no longer allow those affected to be assessed as 'single'. This situation leaves one person to pay both household costs and long-term care facility expenses with less GIS support. The Minister of Families, Children and Social Development announced that the guidelines were to be changed to what they were prior to January 2017.</p> <p><b>Current Situation:</b> After a policy change, a number of applicants were denied GIS. The change in policy was reversed and analysts would need to go through all ITRDS notes to identify the clients that were denied.</p> <p><b>Solution Provided:</b> The CDO built a model to identify those automatically, saving numerous FTEs that can be allocated to other tasks.</p>
<p><b>Innovation, Information and Technology Branch (IITB), Shared Service Canada (SSC) (Data Lake)</b></p> <p><i>(Project Underway)</i></p>	<p><b>Current Situation:</b> ESDC analysts and external researchers requiring protected data must go through a lengthy and complicated process to access data.</p> <p><b>Solution Proposed:</b> Exploring a Protected B Cloud Pilot on a variety of data sources including image and speech to provide easier access and advanced analytical tools to researchers.</p>



## Annex C : Description of projects by stakeholders (7 of 8)

Stakeholders	Projects
<p><b>Public Affairs and Stakeholder Relations Branch (PASRB) – Media Monitoring</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b>          Everyday, thousands of articles and news stories on ESDC or its Ministers are produced. Keeping track of what is being said about the department, its programs or its Ministers with manual techniques is becoming more and more difficult as the speed at which news are produced.</p> <p><b>Current Situation:</b>          The Government of Canada (GoC), through Public Safety Canada, purchases a list of articles pertaining to all federal departments. Those lists are then shared with departments for further analysis. ESDC employees currently come in to work at 3 A.M. in order to open and read between 800 and 1000 articles. Of those, about 100 are evaluated and summarized for ESDC's higher management. Manual monitoring of a sample of all news related to the department is resource intensive and time consuming.</p> <p><b>Solution Proposed:</b>          The CDO is working on a tool that will rate not only a sample, but all the articles provided by Public Safety Canada everyday. It will rate articles according to their content, the mention of an ESDC program or if ESDC's Ministers are mentioned. Once the monitoring part of the model is up and running, an automatic summarizer will be implemented.</p>

## Annex C : Description of projects by stakeholders (8 of 8)

Stakeholders	Projects
<p><b>Citizen Service Branch(CSB) – MSCA/MSBA Operations team</b></p> <p><i>(Project Completed)</i></p>	<p><b>Business Context:</b>            My Service Canada Account (MSCA) is an online portal administered by Service Canada, providing convenient and secure access to information regarding: EI, CPP, OAS            MSCA has an online feedback form where users can submit suggestions on how to improve MSCA by filling out this form. The feedback form comments provide valuable information to help improve the MSCA services.</p> <p><b>Current Situation:</b>            MSCA analysts spend up to <i>25 hours</i> on a monthly basis manually analyzing client feedback. Categorization and analysis is an important task, but the manual process takes away time from MSCA team analysts conducting more in-depth work.</p> <p><b>Solution Proposed:</b>            The research project conducted by the Service Research Division (SRD) was successful in applying machine learning techniques to MSCA feedback form comments in order to facilitate automation of comment categorization. The process of user-feedback categorization now takes a few minutes instead of hours. The time- saving results in more resources being devoted to addressing user concerns and improving government services. SRD also developed an intuitive graphical user interface to interact with the model and is currently investigating visual and report-generating tools to better analyze MSCA user feedback.</p>

## Annex D : Objective of a Data Strategy

- The main objective of the data strategy is to get data into the hands of people who can **drive value** with the work that they do.
- There are 6 work streams that will make that happen in a secure way, that respects the privacy of individuals, **proving that data can be both more secure and more accessible**.
- Two work streams in particular Data Access and Data Science, will enable ESDC employees and partners such as members of ESDC the Canadian Research Data Centre Network, to perform analytics and research that **will drive both our policy and service mandates**.





## Annex E: Status: Progress on leveraging our data assets...

### Access

- Enhancing collaboration with **Statistics Canada, academia, and other departments**:
  - Strengthened joint ESDC-StatCan governance (e.g., ADM-ACS, DM-CS)
  - ESDC-StatCan Data Sharing Roadmap approved and implementation underway
  - Created Interdept. Working Group on Children's Data to increase data sharing and address gaps
- Public release of over 100 data sets as **open data**
- Development and early implementation of a **Hackathon Strategy** (e.g., on-track for a poverty reduction hackathon Spring 2018)
- Working with Justice Canada and other stakeholders to identify barriers and provide recommendations to enable data sharing through **Privacy Act Reform**
- Chief Data Officer co-chairing new **Privacy and Data Access Committee**

### Data Science

- Building and piloting new approaches to leveraging our data
  - **Leveraging unstructured data** (e.g., eliminated 50,000 pieces of work previously done manually by revamping the returned T4 slips process, automated email triage of Old Age Security clients, thematic analysis of online Poverty Reduction Strategy Consultation)
  - **Supporting proactive decisions** (e.g., Predicting which client files require additional documentation at receipt to reduce delays and support better decision-making by CPP-D agents to eliminate costly appeals of decisions)
  - **Preparing for future technologies** (e.g., developing Analytics Program and Artificial Intelligence strategy, exploring Speech Recognition technology to identify how it can be used for client identification and analytics)

See Annex A for detailed overview of progress milestones





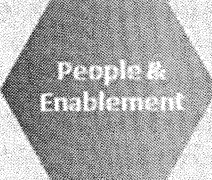
Employment and  
Social Development Canada

Emploi et  
Développement social Canada

Canada



## ... while building key data foundations

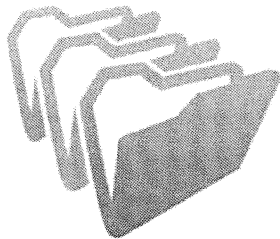
	<ul style="list-style-type: none"> <li>• Piloting formalized <b>data stewardship</b> function with HR to demonstrate the value of a systematic process to identify and resolve data issues.</li> <li>• Piloting improved tools and methodology for assessing <b>data quality</b> to improve quality of performance information reporting.</li> <li>• Develop performance metrics proposal to articulate the anticipated outcomes from enhanced data governance.</li> </ul>
	<ul style="list-style-type: none"> <li>• <b>Enterprise Business Information and Analytics Reference Architecture</b> proposal developed with Innovation, Information and Technology Branch (IITB) and approved by the ESDC Enterprise Architecture Review Board (EARB)</li> <li>• Chief Data Office leading Service Transformation Plan <b>Single Client View “pod”</b></li> </ul>
	<ul style="list-style-type: none"> <li>• <b>Post-Secondary Recruitment</b> campaign for the data and research streams likely to result in a much larger pool of high quality candidates at the masters/PhD level with 120 advancing to the interview phase (compared to a pool of 15 in 2017).</li> <li>• <b>Collaborating with GoC stakeholders</b> to advance data initiatives (e.g. DG Data Leads Committee, Metadata Standards Sub-group, Open Government Multi-Stakeholder Forum, hosting 2018 GoC Data Conference)</li> <li>• Developing <b>Communications &amp; Engagement and Data Literacy</b> plans</li> </ul>

See Annex A for detailed overview of progress milestones

Canada

## Annex F : To What Purpose is ESDC using Data Science?

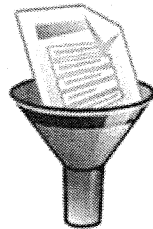
### Categorization



Dividing the data into groups, categories, or topics that are related

- Triaging work
- Segmenting clients
- Identifying topics or themes

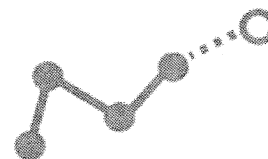
### Key Information Extraction



Isolating important information and extracting the useful part

- Chatbots\*
- Autofill forms or templates
- Descriptive information

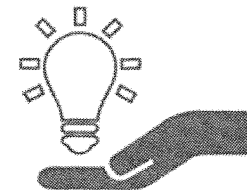
### Prediction



Using historical relationships to predict future outcomes

- Advise agents on decision to be taken
- Predict future workload
- Predict potential benefits for clients

### Technical Expertise to Advise



Exploring emerging technologies and approaches to inform decisions

- Help ESDC sign better contracts
- Identify key emerging tech and adapt it for ESDC's needs.
- Repurpose and adapt models for quick wins

\* See glossary in annex A

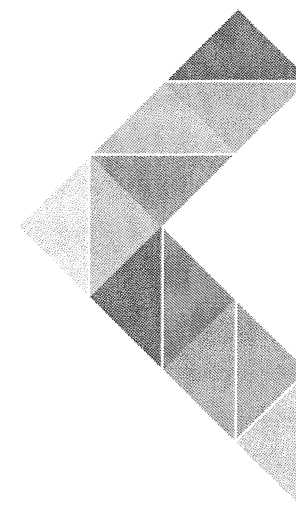
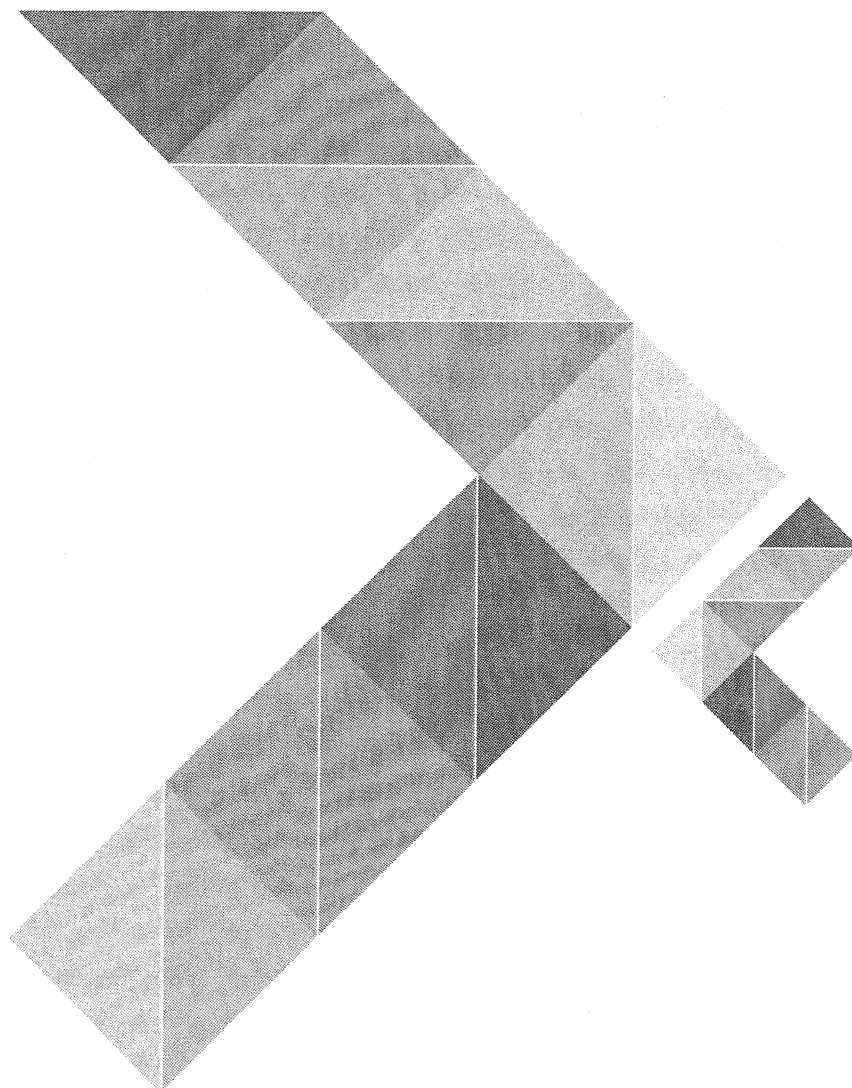




Employment and  
Social Development Canada

Emploi et  
Développement social Canada

Service  
Canada



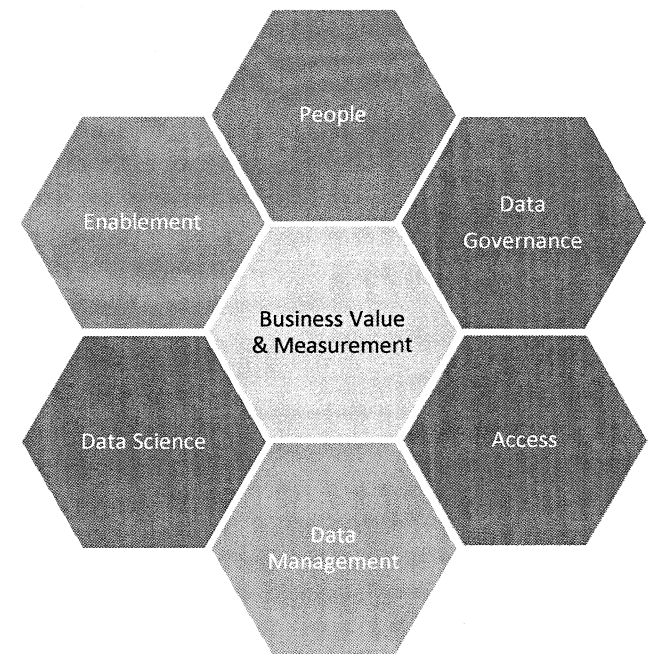
# Data Science

Chief Data Office, SSPB

# Data Science

## Purpose

- Data science is one of the streams of the Data Strategy that aims to provide people with rapid, secure and authorized access to quality data in a way that respects personal privacy and delivers value by giving them the skills, tools, and processes needed to maximize the impact of our enterprise data asset.
- Data science is about using analytic techniques, such as machine learning\*, sentiment analysis and natural language processing, on data to solve business problems.
  - It mines large amounts of data at a granular level to identify complex behaviors, patterns and trends that uncover hidden insights that enable organizations to make smarter decisions.
  - The CDO has led a number of data science pilot projects that are described in slides 10 to 20.



Data Strategy Elements

# Data Science

## Why does ESDC need Data Science?



- ESDC is a heavily manual organization
  - Current methods to process work for clients or generate reports are slow, inefficient, inconsistent, and prone to error
- ESDC has a huge amount of unstructured data (text, scans, audio)
  - A majority of the information held by ESDC is not used because it is hard to access and process using traditional analytical tools
- ESDC's service delivery is too often reactive and not proactive
  - The department is continuously playing catch-up on existing workload inventories. There are millions of items in the backlog and low value items often crowd the critical work. This limits the resources we can put towards more proactive service delivery strategies

By putting data science in place throughout the organization, the CDO Office is enabling ESDC analysts to deliver better results more efficiently, add greater value in their work and provide better service to Canadians.

# Data Science

## How is the CDO Office putting Data Science into place?

- Leverage analytics to maximize the value of data
  - Work with partners on pilot projects to deliver immediate value, while demonstrating future potential
- Develop governance for analytics
  - Provide guidance to the organization on the skills, tools, techniques and processes needed for analytics.
- Democratize analytics
  - Educate on the role and potential of analytics so that groups across the organization can do it independently.
- Provide Expert Advice
  - Support branches in negotiations with external vendors to get the best possible contracts by making sure the department retains all IP from developed models and receives maximum value for its investment.
- Partner with the Innovation, Information and Technology Branch (IITB) to upscale proven data science pilots.
  - Transition pilots into business solutions that integrate Data Science into how we work.



# Where are we right now ?

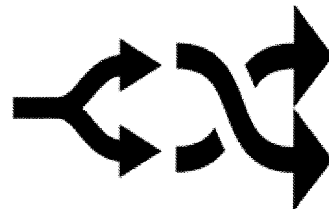


## « Replicate human actions »

Used for rules based,  
simple to complex process  
Faster handling time  
Higher volumes  
Reduced errors and costs

### ESDC Current Production

- Smart Recommender System
- Workflow triage
- Supported Information Retrieval



## « Augment human judgement »

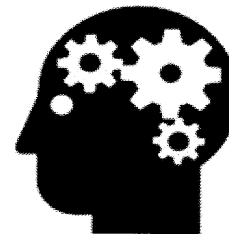
- Used for judgement  
based processing  
- Machine learning  
capability  
- Interpret human behavior

### ESDC's Pilots

- SCARLET
- T4 Investigating
- Labour question answering machine
- Homelessness and Poverty survey analysis
- Involuntary Separation

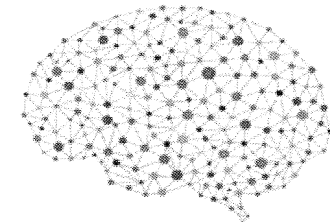
### ESDC's Potential

- Client Decision Prediction
- Question answering (Chatbot)
- Speech to text and voice recognition
- Document Drafting



## « Augment human intelligence »

Used for predictive  
« Decisioning »  
Dynamically self-adaptable  
and managing



## « Mimics human intelligence »

Acquire human-like  
thought processing and  
capabilities

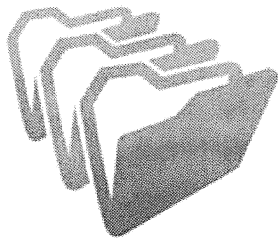
### Where the current big players are:

- |  |  |
|--|--|
| <ul style="list-style-type: none"> <li>• Netflix and Amazon recommendation tools</li> <li>• Tesla's autopilot</li> </ul> | <ul style="list-style-type: none"> <li>• SpaceX rocket landing algorithms</li> <li>• AlphaGo Zero</li> <li>• IBM Watson; Apple devices; Google A.I.</li> </ul> |
|--|--|

# Data Science

## What is the CDO Office using Data Science to do?

### Categorization



Dividing the data into groups, categories, or topics that are related

- Triaging work
- Segmenting clients
- Identifying topics or themes

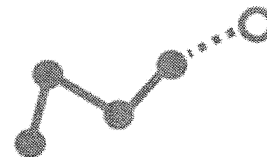
### Key Information Extraction



Isolating important information and extracting the useful part

- Chatbots\*
- Autofill forms or templates
- Descriptive information

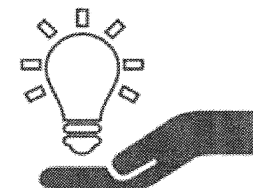
### Prediction



Using historical relationships to predict future outcomes

- Advise agents on decision to be taken
- Predict future workload
- Predict potential benefits for clients

### Technical Expertise to Advise



Exploring emerging technologies and approaches to inform decisions

- Help ESDC sign better contracts
- Identify key emerging tech and adapt it for ESDC's needs.
- Repurpose and adapt models for quick wins

\* See glossary in annex B



# Data Science

## How do machines learn?

- The idea behind machine learning is to train machines the way someone would train a human – through experience and expert guidance
- Supervised learning involves an analyst manually coding examples (a training set) that lets the machine develop an algorithm to mirror those decisions
  - For example, assign a number of documents to group 1, 2, or 3 that allows the machine to see a pattern that it uses to classify future documents
- Unsupervised learning involves the machine being given a defined task without human assisted examples of the “right” and “wrong” answers
  - For example, divide these documents into 4 groups without specifying the groups, based on whatever the computer sees as relevant
- Reinforcement learning is where the machine is able to interact with its environment to determine whether or not it achieved its desired outcome and then adjusts its decision making to improve
  - For example, when Netflix recommends a movie and then the user stops watching after 5 minutes, the algorithm learns that that wasn't a good recommendation and refines its algorithm



# Data Science

## Key CDO Data Science focus: Artificial Intelligence

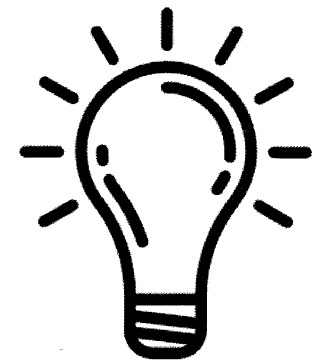
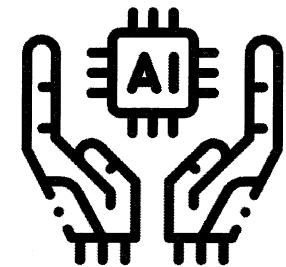
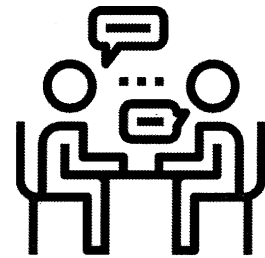
- Leading the Department on A.I.
  - We are developing an **A.I.\* Strategy** for the Department that outlines where we need to go and how we'll get there.
- Building a Departmental A.I. suite
  - In the first project, the CDO is working with the Transformation and Integrated Service Management Branch (TISMB) to implement a workflow triage A.I. that is expected to avoid the department \$500,000 in costs annually.
- Establish A.I. Thought Leadership
  - Support partner branches in negotiations with external vendors, make sure ESDC retains all IP and can properly value the services provided.
  - Working with partner branches to see how A.I. can be beneficial to their work.

\* See glossary in annex B

# Data Science

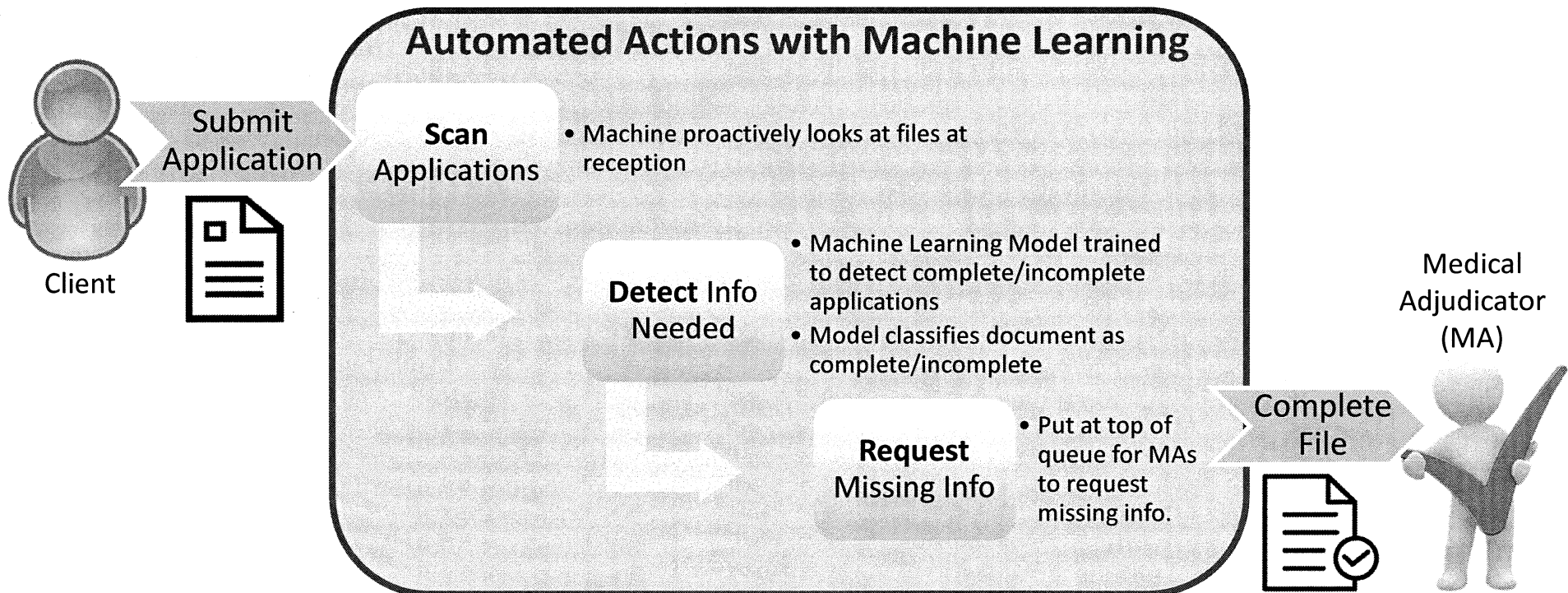
## Data Science Projects

- The CDO is leading numerous pilot projects with multiple branches in the department.
- The following slides describe a few examples of solutions provided by the CDO that are:
  - Enabling proactive decisions
  - Automating manual processes
  - Leveraging unstructured data (text, image, audio)
  - Preparing for future technologies



# Enabling proactive decisions

Business Context	The Canada Pension Plan (CPP) provides disability benefits to people who are disabled and cannot work at any job on a regular basis through the CPPD program. Medical adjudicators (MA) determine eligibility after reviewing detailed applications involving significant medical documentation
Problem	Applicants have to wait for a MA review to determine if they require additional documentation. Often the MA will not look at an application for several months, making it difficult to meet the service standard of 120 days when extra information is needed.
Value	<ul style="list-style-type: none"> <li>➤ Clients get their benefits faster</li> <li>➤ Fewer rejections and appeals due to missing information</li> <li>➤ MA's time more efficiently used to evaluate complete files</li> </ul>



9

# Automating Manual Processes

Business  
Context

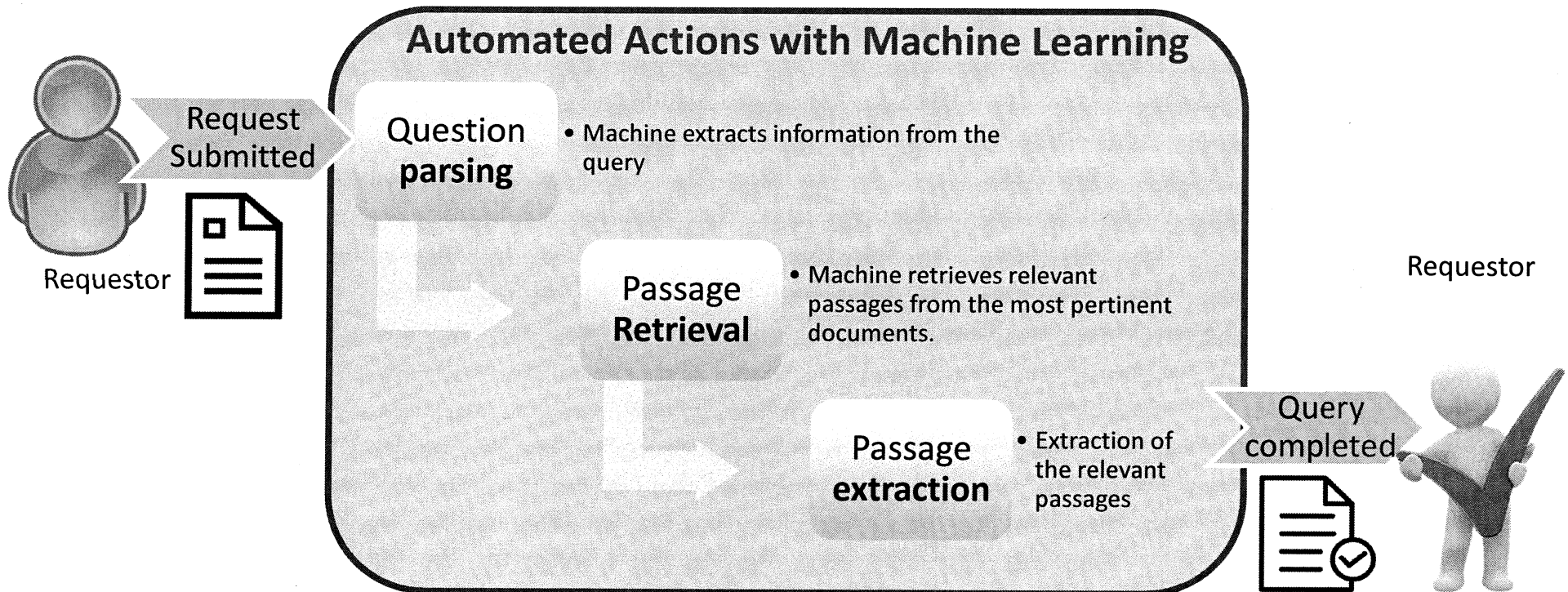
The labour program receives questions regarding collective agreements related to federally regulated occupations. Analysts need to extract information manually from collective agreements by opening documents and searching for keywords.

Problem

Due to the large number of collective agreements (over 50,000 agreements) analysts can only look at a sample of documents. Manually going through numerous document is not efficient, difficult to reproduce, inconsistent and time consuming.

Value

- Faster querying system
- Research is done on all the collective agreements, not a sample
- Model can be reuse on other similar question-answer problems





## Automating Manual Processes (2)

### Business Context

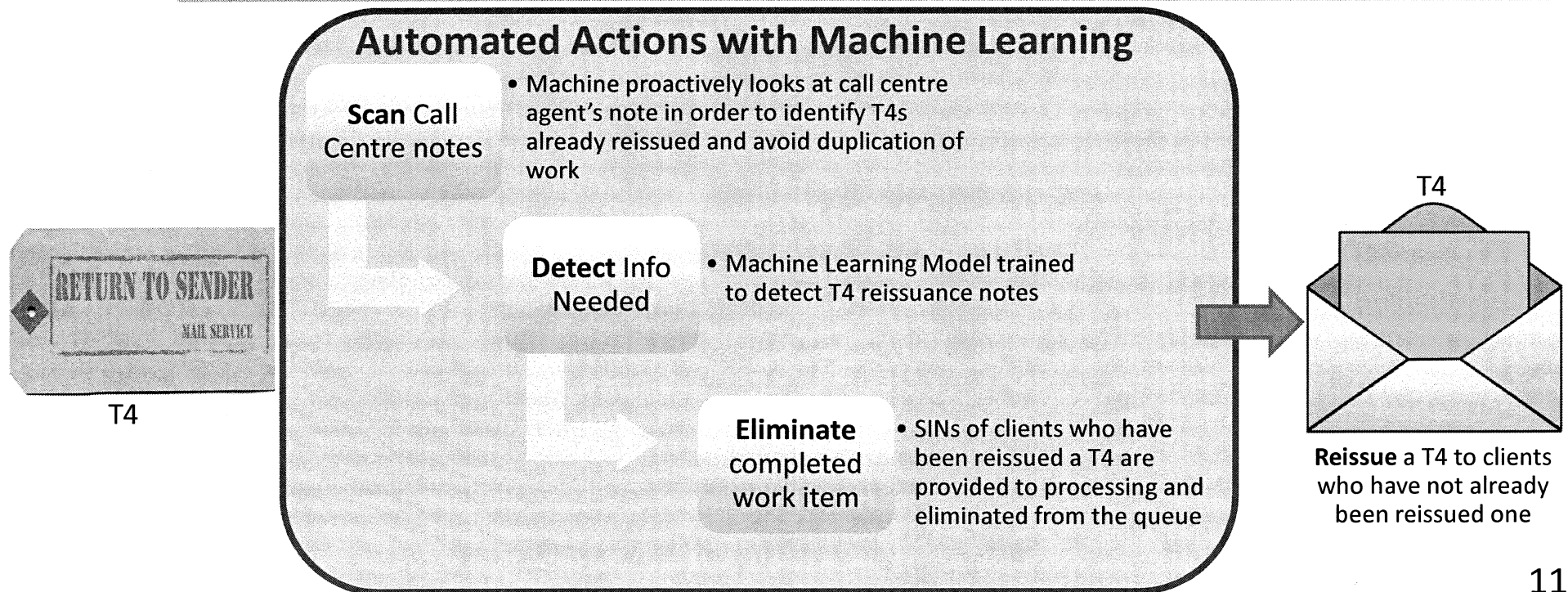
Each year Canada Pension Plan (CPP) and Old Age Security (OAS) recipients receive a T4 form containing the benefits received that they will need to report on their tax return. If they are recipients of the CPP the slip is a T4A(P) while it's a T4A (OAS) for OAS.

### Problem

Each year many T4s are returned to the Service Canada processing network due to address changes or other reasons. A large number of clients previously followed-up with Service Canada to request a duplicate T4, which is immediately reissued. Processing still performs resource intensive investigation of each returned item individually.

### Value

- Low-value work eliminated: 50,000 work items, or 2.5 FTEs, saved annually, while development only used 0.1 FTE
- Work inventory is reduced, with faster processing for remaining clients
- Once built, the models can be repurposed to answer new questions



# Leveraging unstructured data (text, images, audio)

## Business Context

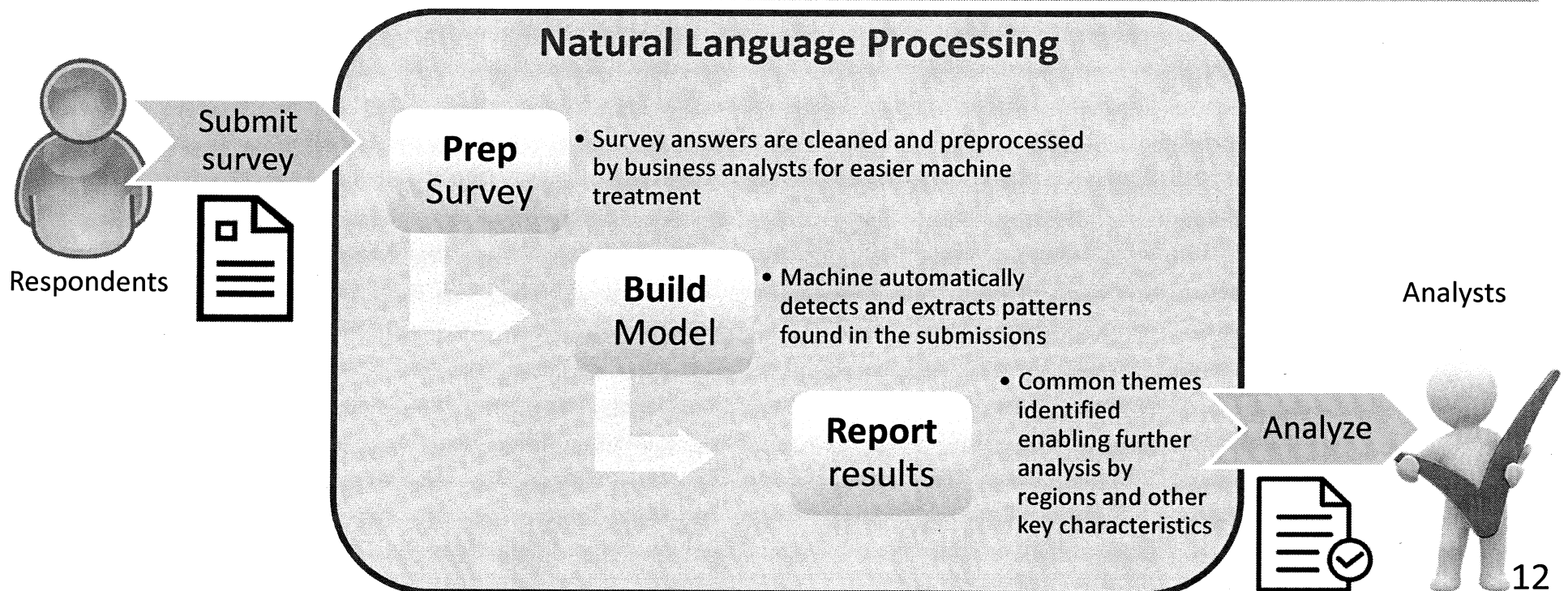
ESDC runs numerous consultations that contain at least one open ended section where participants can express themselves on a specific topic. In 2017, ESDC ran a consultation around the theme of poverty reduction. As part of this consultation, the department received over 800 submissions in free text (emails, online forum and story submissions).

## Problem

To get insight into the Poverty Reduction Strategy, analysts would read through different types of submissions (emails, online forums and surveys) from over 800 respondents and try to extract key information manually. This method is resource intensive and inconsistent.

## Value

- Fast, unbiased and reproducible insights (took one student 3 weeks to build a model of themes)
- Scalable to very large data sets
- Analysts can delve into explaining results rather than identifying trends saving about 1/6 FTE that would have been allocated to reading the submissions.



# Preparing for future technologies

## Business Context

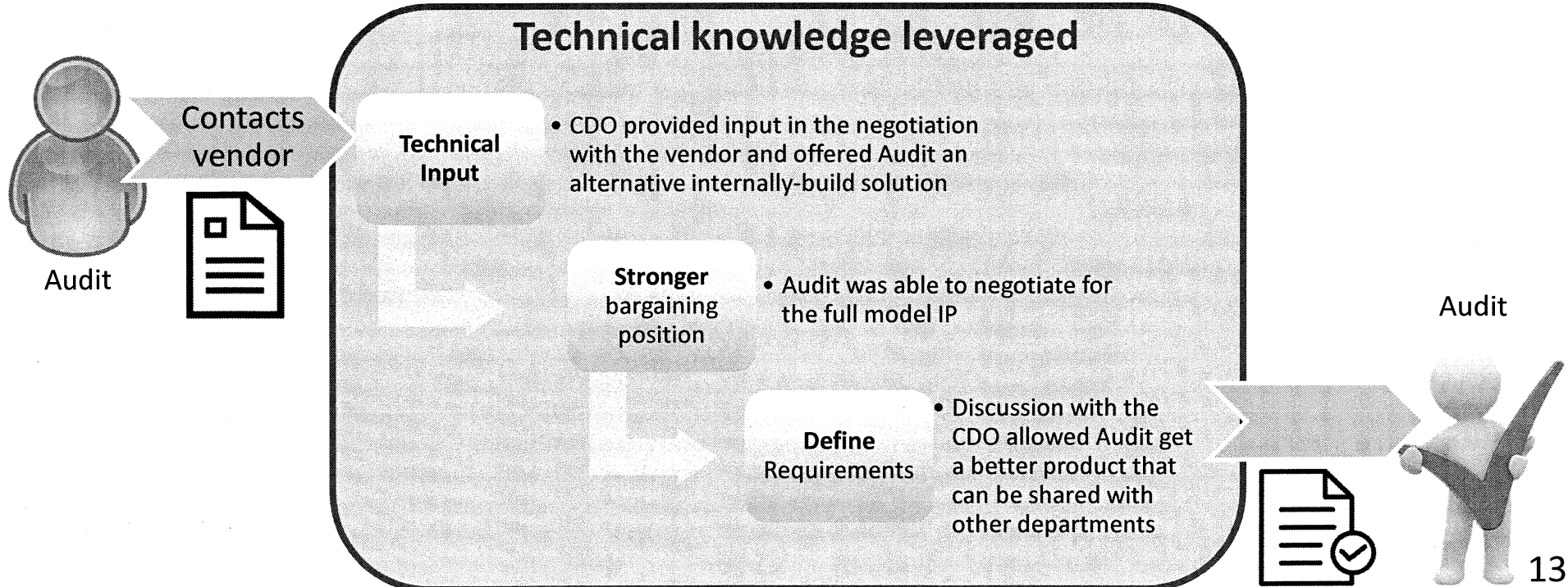
External Artificial Intelligence (AI) tools are increasingly being considered by various branches in their attempt to leverage their data assets. However, the knowledge of the different techniques in the department is sparse so leveraging external vendors is a common option explored.

## Problem

Audit contacted a private vendor to acquire a machine learning categorization tool. The private vendor would have required the department to pay development costs up front, acquire a subscription to use the model and ESDC would own no intellectual property (IP).

## Value

- IP rests with ESDC allowing the model to be repurposed, reused, and shared with other Departments
- Audit contracted for a better product at the same costs
- External expertise is leveraged to hit tight timelines





# Data Science

## Next Steps

- Put models into **production** to deliver value
  - Working closely with the Transformation and Integrated Service Management Branch operations
- Support **Service Transformation Plan** activities
  - Regular participation in the acceleration hub and advising on specific initiatives
- Deliver the first stage of a draft **Analytics Program** in Fall 2017
  - Will be used to engage stakeholders across the organization
- Finalize the **Artificial Intelligence Strategy**
- Work with partners across the organization to discover where data science can help them and **build organizational capacity**
  - Leading data and research streams of 2017-18 post-secondary recruitment to identify strong technical candidates for all of ESDC
- Work with the Innovation, Information and Technology Branch to make **AI sustainable**
  - Developing long-term solution for the models, how they are stored, accessed and maintained

## ANNEX A: Pilot Project Details

Continued

Stakeholders	Projects
<b>Human Resources Service Branch</b>  <i>(Project Completed)</i>	<p><b>Business Context:</b>            In order to improve employee retention and identify shortcomings in employer-employee relations, HRSB runs an Employee Exit Survey to all employees leaving the department. Some of the fields in this survey are free text.</p> <p><b>Current situation:</b>            Extraction of key information would need to be done manually, which is inconsistent and resources intensive.</p> <p><b>Solution Provided:</b>            Automated the analysis of the employee exit survey using topic modelling. Common themes that may have otherwise fallen through the cracks were identified (i.e. the link between mental health and housing). Using natural language processing is faster, unbiased, scalable and reusable.</p>
<b>Internal Audit Service Branch (IASB)</b>  <i>(Project Completed)</i>	<p><b>Business Context:</b>            External Artificial Intelligence (AI) tools are increasingly being considered by various branches in their attempt to leverage their data assets. However, the knowledge of the different techniques in the department is sparse so leveraging external vendors is a common option explored.</p> <p><b>Current Situation:</b>            Audit contacted a private vendor to acquire a machine learning categorization tool. The private vendor would have required the department to pay development costs up front, acquire a subscription to use the model and ESDC would own no intellectual property (IP).</p> <p><b>Solution Provided:</b>            CDO provided input in the negotiation with the vendor and offered Audit an alternative internally-build solution. Audit was able to negotiate for the full model IP with no ongoing subscriptions fees. The discussion between the CDO and Audit allowed Audit to better define their requirements.</p>

Stakeholders	Projects
<p><b>Income Security and Social Development Branch (ISSD) (CPPD)</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b>            Canada Pension Plan (CPP) provides disability benefits to people who have made enough contributions to the CPP and who are disabled and cannot work at any job on a regular basis. Medical adjudicators (MA) determine, after reviewing your application, your eligibility.</p> <p><b>Current Situation:</b>            CPPD applications have to wait for a medical adjudicator's (MA) review to determine if they require additional documentation. Often this review takes place after several months making it hard to meet the service standard of 120 days.</p> <p><b>Solution Provided:</b>            Predicting which client files require additional documentation at receipt to reduce delays and support better and more consistent decisions by CPP-D agents to eliminate costly appeals of decisions.</p>
<p><b>Income Security and Social Development Branch (ISSD) (Homelessness Strategy)</b></p> <p><i>(Project Completed)</i></p>	<p><b>Business Context:</b>            ESDC runs, year after year, numerous consultations which almost always contain at least one open ended section where participants can express themselves on a specific topic. In 2017, ESDC ran a consultation around the theme Homelessness. As part of this consultation, the department received hundreds of submissions in free text.</p> <p><b>Current Situation:</b>            To get insight in the Homelessness Strategy, analysts would read through all submissions and try to extract key information manually. This method is resource intensive, inconsistent and not reproducible.</p> <p><b>Solution Provided:</b>            Through topic modelling, the CDO provided with a fast, scalable and efficient way to extract key information.</p>

Stakeholders	Projects
<p><b>Labour Program – Information Retrieval System</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b>            About 1 million employees in Canada work in federally regulated occupations. The Labour Program promotes cooperation and fairness and provides expert advice and assistance on labour relations matters to workplaces within the federal jurisdiction. To that intent, the Labour Program has developed numerous services, measures and initiatives to assist employers and employees in creating and maintaining a workplace that is conducive to good industrial relations.</p> <p><b>Current Situation:</b>            The labour program, receives requests on collective agreements related to federal jurisdiction occupations, but most extract information manually from a sample of collective agreements because of the large number of agreements.</p> <p><b>Solution Provided:</b>            Creating a question and answer tool to categorize and facilitate responses to queries of collective agreements. Model allows for a faster, more efficient and reproducible way of querying collective agreements. The model required 0.2 FTE for its creation and can be reused at almost no cost.</p>
<p><b>Legal Services (Justice Canada)</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b>            External A.I. tools are more frequent considered by different branches in their attempt to leverage their data assets. Legal is facing an ever growing volume of documentation and needs to keep up with the private sector which is using those tools. However, the knowledge of the different techniques in the department is sparse.</p> <p><b>Current Situation:</b>            Legal is interested in acquiring an A.I. solution to keep up with the ever increasing amount of documentation. Legal lacks the proper knowledge to define their requirements in terms of A.I. to external vendors.</p> <p><b>Solution Proposed:</b>            Providing technical expertise to advise on Artificial Intelligence projects with external contractors to facilitate the acquisition of the right solution.</p>



Stakeholders	Projects
<p><b>Strategic and Service Policy Branch (SSPB) – Poverty Reduction Strategy</b></p> <p><i>(Project Completed)</i></p>	<p><b>Business Context:</b> ESDC runs, year after year, numerous consultations which almost always contain at least one open ended section where participants can express themselves on a specific topic. In 2017, ESDC ran a consultation around the theme Poverty Reduction. As part of this consultation, the department received hundreds of submissions in free text.</p> <p><b>Current Situation:</b> To get insight in the Poverty Reduction Strategy, analysts would read through all submissions and try to extract key information manually. This method is resource intensive, inconsistent and not reproducible.</p> <p><b>Solution Provided:</b> Automated the extraction of key themes from hundreds of online submissions related to the Poverty Reduction Strategy.</p>
<p><b>Service Transformation Plan (STP)</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b> ESDC is in the midst of transforming the way we deliver services to Canadians. Among the transformation envisioned are modern authentication methods and automated dialoguing solutions.</p> <p><b>Solution Envisioned:</b> Exploring speech recognition technology to identify how it can be used for both client identification as well as analytics.</p>
<p><b>Transformation and Integrated Service Management Branch (TISMB) – OAS related mail</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b> ESDC is in the midst of transforming the way we deliver services to Canadians. Among the transformation envisioned are modern authentication methods and automated dialoguing solutions.</p> <p><b>Current Situation:</b> A significant amount of emails relating to OAS are hitting the system. Many of those are related to work that cannot be actioned or duplicate work that is currently underway or has been addressed. The current email triaging requires many hours of manual effort, which could otherwise be used in more important areas of processing.</p> <p><b>Solution Provided:</b> Through a number of techniques, the CDO streamlined, closed and aggregated email workflow by triaging emails as they arrived.</p>

Stakeholders	Projects
<p><b>Transformation and Integrated Service Management Branch – OAS overpayment</b></p> <p><i>(Project Completed)</i></p>	<p><b>Business Context:</b>  Occasionally under/overpayments are issued to Canadians and Service Canada is notified by the clients. The errors are noted in a system. In order to ensure that every Canadians gets what he/she is entitled too, intensive manual efforts are needed.</p> <p><b>Current Situation:</b>  In a number of cases, overpayments and underpayments are issued by ESDC to clients. Information Technology Renewal Delivery System (ITRDS) notes contains information on those over/underpayments. The Transformation and Integrated Service Management Branch employees are manually going through files and extract the information from the ITRDS Notes. This work is time consuming, resource intensive and inconsistent.</p> <p><b>Solution Provided:</b>  The CDO used predictive models to flag ITRDS notes reporting an over/underpayment amount over a certain threshold. The model demonstrated accurately that it could extract specific information. The model build by the CDO using only 1/3 FTE is efficient, consistent, is replicable and frees up valuable analysts time. Also, the model developed here can be repurposed to do similar task internally at ESDC or in other departments facing similar issues.</p>
<p><b>Transformation and Integrated Service Management Branch – T4 Processing</b></p> <p><i>(Project Completed)</i></p>	<p><b>Business Context:</b>  Each year Canada Pension Plan (CPP) and Old Age Security (OAS) recipients receive an information slip containing the information they will need to report on their tax return. If they are recipients of the CPP the slip is a T4A(P) while it's a T4A (OAS) for OAS.</p> <p><b>Current Situation:</b>  Service Canada processing network receives numerous returned T4's due to changes in client's address or for other reasons. A significant number of clients follow-up with Service Canada to request a duplicate tax slip. Processing is not made aware which in return create intensive manual investigation.</p> <p><b>Solution Provided:</b>  The model identifies when a client's T4 as being reissued, provided the SIN associated with that T4 to processing and eliminates it from the queue. 50,000 work items, or 2.5 FTEs, saved annually, while development only used 0.1 FTE. Work inventory is reduced, with faster processing for remaining clients. Once built, the models can be repurposed to answer new questions.</p>

Stakeholders	Projects
<p><b>Transformation and Integrated Service Management Branch – EI Application</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b> ESDC is still relying heavily on paper applications for the numerous programs in its portfolio. Among them, the Employment Insurance (EI) program requires Canadians to fill a paper application.</p> <p><b>Current Situation:</b> EI applications are manually scanned and tags are manually assigned to categorized them.</p> <p><b>Solution Proposed:</b> Automating the classification of EI applications allowing the reallocation of valuable employee time.</p>
<p><b>Transformation and Integrated Service Management Branch – Involuntary Separation – GIS</b></p> <p><i>(Project Completed)</i></p>	<p><b>Business Context:</b> The GIS provides a monthly non-taxable benefit to Old Age Security (OAS) pension recipients who have low income and are living in Canada. In January 2017, guidelines for Guaranteed Income Supplement (GIS) were changed. The GIS guidelines, in the case of a partner admitted to long-term care facility would no longer allow those affected to be assessed as 'single'. This situation leaves one person to pay both household costs and long-term care facility expenses with less GIS support. The Minister of Families, Children and Social Development announced that the guidelines were to be changed to what they were prior to January 2017.</p> <p><b>Current Situation:</b> After a policy change, a number of applicants were denied GIS. The change in policy was reversed and analysts would need to go through all . Information Technology Renewal Delivery System (ITRDS) notes to identify the clients that were denied.</p> <p><b>Solution Provided:</b> The CDO built a model to identify those automatically, saving numerous FTEs that can be allocated to other tasks.</p>
<p><b>Innovation, Information and Technology Branch (IITB), Shared Service Canada (SSC) (Data Lake)</b></p> <p><i>(Project Underway)</i></p>	<p><b>Current Situation:</b> ESDC analysts and external researchers requiring protected data must go through a lengthy and complicated process to access data.</p> <p><b>Solution Proposed:</b> Exploring a Protected B Cloud Pilot on a variety of data sources including image and speech to provide easier access and advanced analytical tools to researchers.</p>



Stakeholders	Projects
<p><b>Public Affairs and Stakeholder Relations Branch (PASRB) – Media Monitoring</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b>            Everyday, thousands of articles and news stories on ESDC or its Ministers are produced. Keeping track of what is being said about the department, its programs or its Ministers with manual techniques is becoming more and more difficult given the speed at which news is produced.</p> <p><b>Current Situation:</b>            The Government of Canada (GoC), through Public Safety Canada, purchases a list of articles pertaining to all federal departments. Those lists are then shared with departments for further analysis. ESDC employees currently come in to work at 3 A.M. in order to open and read between 800 and 1000 articles. Of those, about 100 are evaluated and summarized for ESDC's higher management. Manual monitoring of a sample of all news related to the department is resource intensive and time consuming.</p> <p><b>Solution Proposed:</b>            The CDO is working on a tool that will rate not only a sample, but all the articles provided by Public Safety Canada everyday. It will rate articles according to their content, the mention of an ESDC program or if a Minister is mentioned. Once the monitoring part of the model is up and running, an automatic summarizer will be implemented.</p>



NOW AND TOMORROW EXCELLENCE IN EVERYTHING WE DO



2018 SC-CAMS 000049

# Data Science SAC Meeting

Chief Data Office, SSPB



Employment and  
Social Development Canada

Emploi et  
Développement social Canada

Canada

# Key Themes

- Data science is a core element of ESDC's Data Strategy and we believe we can take greater steps in using AI to explore new areas of client service, as opposed to simply using it to make our current quality of service more efficient.
- We have had a lot of success generating value from early analytics/AI pilots, and we need to figure out optimal processes to integrate those models into production, including infrastructure best practices and governance around the process. To what degree should rules and guidelines around decisions made by AI models be developed in tandem with Treasury Board?
- Strategic decisions on what to build vs. buy need to be made
- Excitement about the opportunities is high, and there is a lot of buy-in across the department, but ensuring comprehensive culture change is something we'll need to learn to manage



# The world of data has changed, and it means change for ESDC

- **Client expectations have changed**

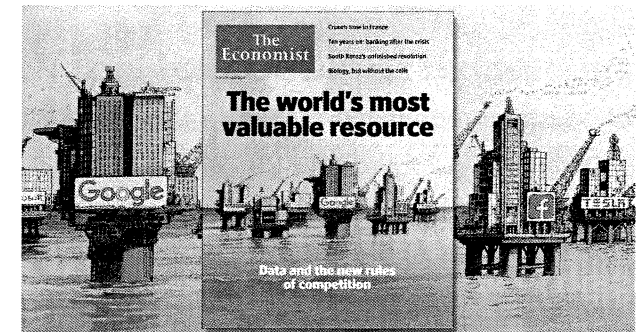
Clients expect private sector level of service, but with a much higher standard for privacy protection. We must integrate our data to support proactive client service within secure and managed environments.

- **ESDC has changed**

Evidence-based decision making and Results & Delivery, the Service Transformation mandate, and a focus on Transparency require access to data. We must know what data we have and how to use it.

- **Technology and Analytical methods have changed**

We have fallen behind in the underlying investments needed to use and extract the value of data. We need people, technology and an analytics program to tie it all together.



**"I believe that government departments and organizations urgently need to turn their attention to this issue. They need to focus on collecting the right data to support their activities, on ensuring that data is well-managed and up-to-date, and on fully using this data not only to inform their core business, but also to support reporting and continuous improvement."**

***2016 Spring Reports of the Auditor General of Canada, Opening Statement, May 3, 2016***



## ESDC's Data Strategy builds the foundations for improved client service and policy through six work streams

Getting data to everyone who needs it quickly and securely while protecting privacy

Access

Data Science

Building a program to develop analytical capacity to uncover new insights from ESDC's data

### Foundations for leveraging data and analytics

Data  
Governance

Knowing what data we have; what we need; where it is; if we can trust it; who makes decisions; and what rules should be applied

Data  
Management

Ensuring infrastructure is in place to securely store and provide users with access to data and with the tools they need to analyse it

People

Recruiting, retaining and engaging with the people with the skills and experience we need

Enablement

Empowering people to use data and to innovate



Data  
Science

## What challenges are addressed by Data Science?

- ESDC is still a heavily manual organization
  - We have made progress, but more changes are needed to make current methods of processing work for clients, generating reports, or finding answers to questions faster, more efficient, more consistent, and less prone to error
- ESDC has a huge amount of unstructured data
  - A majority of the information held by ESDC is not used because it is hard to access and process using traditional analytical tools
- ESDC is reactive not proactive
  - Currently struggling to meet existing and historical demands which prevents us from focusing on what we should be doing to improve

To meet our need for Data Science, ESDC's Chief Data Office is building a program to develop analytical capacity for using methods such as machine learning, Artificial Intelligence, and others to uncover new insights from ESDC's data.



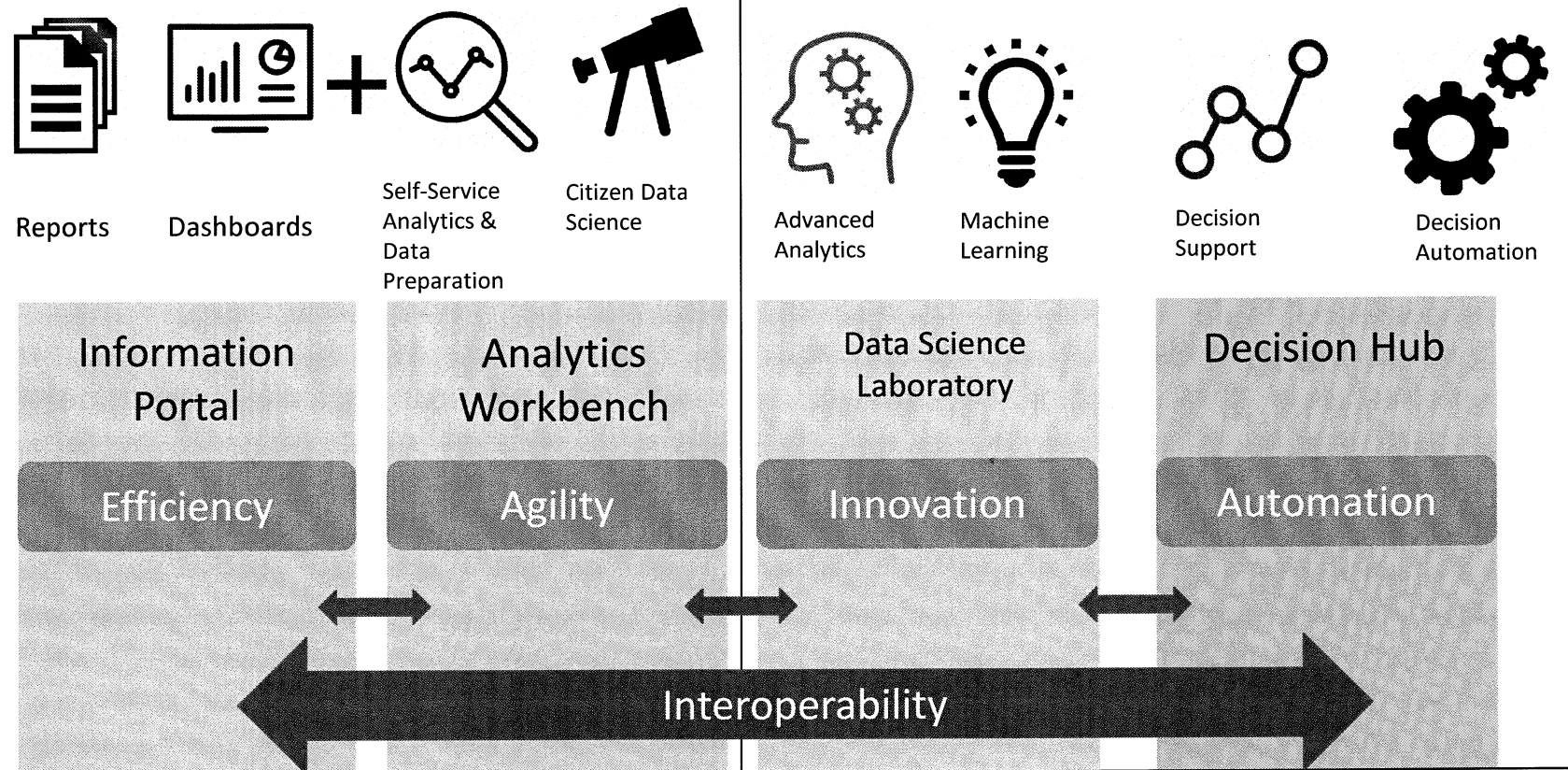


## How is ESDC putting Data Science into place?

- Leverage analytics to maximize the value of data
  - Work with internal partners on pilot projects to deliver immediate value, while demonstrating future potential
- Develop governance for analytics
  - Provide guidance to the organization on the skills, tools, techniques and processes needed for analytics.
- Democratize analytics
  - Educate on the role and potential of analytics so that groups across the organization can do it independently.
- Provide Expert Advice
  - Support branches in negotiations with external vendors to get the best possible contracts by making sure the department retains all IP from developed models and receives maximum value for its investment.
- A partnership between the Chief Data Office and the Innovation, Information and Technology Branch (IITB) to upscale proven data science pilots.
  - Transition pilots into business solutions that integrate Data Science into how we work.



# Spectrum of Analytics





## Key Data Science focus within ESDC's Chief Data Office: Artificial Intelligence

- Leading the Department on A.I.
  - We are developing an **A.I.\* Strategy** for the Department that outlines where we need to go and how we'll get there.
- Building a Departmental A.I. suite
  - In the first project, the CDO is working with the Transformation and Integrated Service Management Branch (TISMB) to implement a workflow triage A.I. that is expected to save the department \$500,000 annually.
- Establish A.I. Thought Leadership
  - Support partner branches in negotiations with external vendors, make sure ESDC retains all IP and can properly value the services provided.
  - Working with partner branches to see how A.I. can be beneficial to their work.

\* See glossary in annex



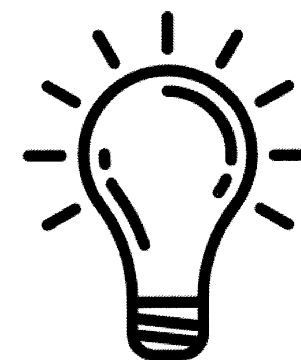
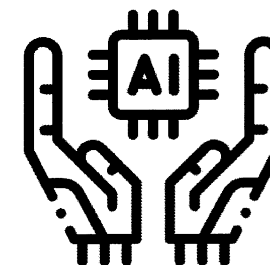
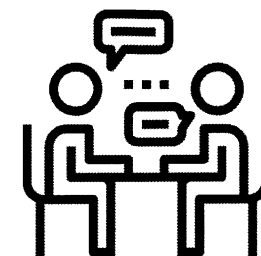
## Questions

- What are the business and IT processes/considerations to deploy pilots into enterprises-wide solutions?
- What is the best way to manage culture change for employees (who see machines taking over their jobs) and clients (potential decreased personal interactions)?
- Have members experienced legal, ethical or other challenges defending decisions that may have been assisted by AI? How was this handled, and what were the outcomes/lessons learned?
- What are the key considerations, risks and benefits of buying/outourcing vs. building internal AI capacity (e.g., IP retention; vendor management, etc.)?



# Data Science Projects

- A number of pilot projects are currently underway with multiple branches in the department.
- The following slides describe solutions provided by the CDO that are:
  - Enabling proactive decisions
  - Automating manual processes
  - Leveraging unstructured data (text, image, audio)
  - Preparing for future technologies
- Detailed summaries of all the pilot projects completed and underway is also provided in an annex



11

## Labour Example

### Business Context

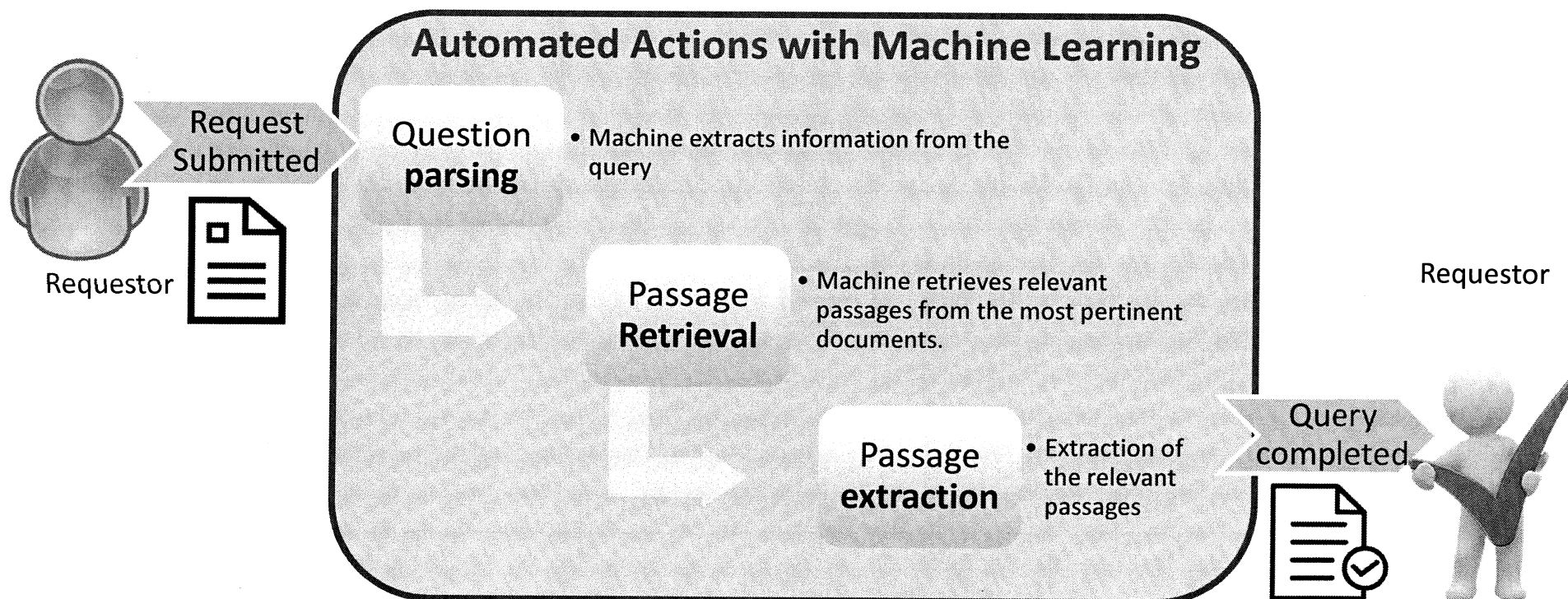
About 1 million employees in Canada work in federally regulated occupations. The Labour Program promotes cooperation and fairness and provides expert advice and assistance on labour relations matters to workplaces within the federal jurisdiction. To that intent, the Labour Program has developed numerous services, measures and initiatives to assist employers and employees in creating and maintaining a workplace that is conducive to good industrial relations.

### Problem

The labour program, upon requests on collective agreements related to federal jurisdiction occupations, need to extract information manually from a sample of collective agreements because of the large number of agreements (+50,000 collective agreements).

### Value

- Faster querying system
- Research is done on all the collective agreements, not a sample
- Model can be reuse on other similar question-answer problems





12

## Enabling proactive decisions

### Business Context

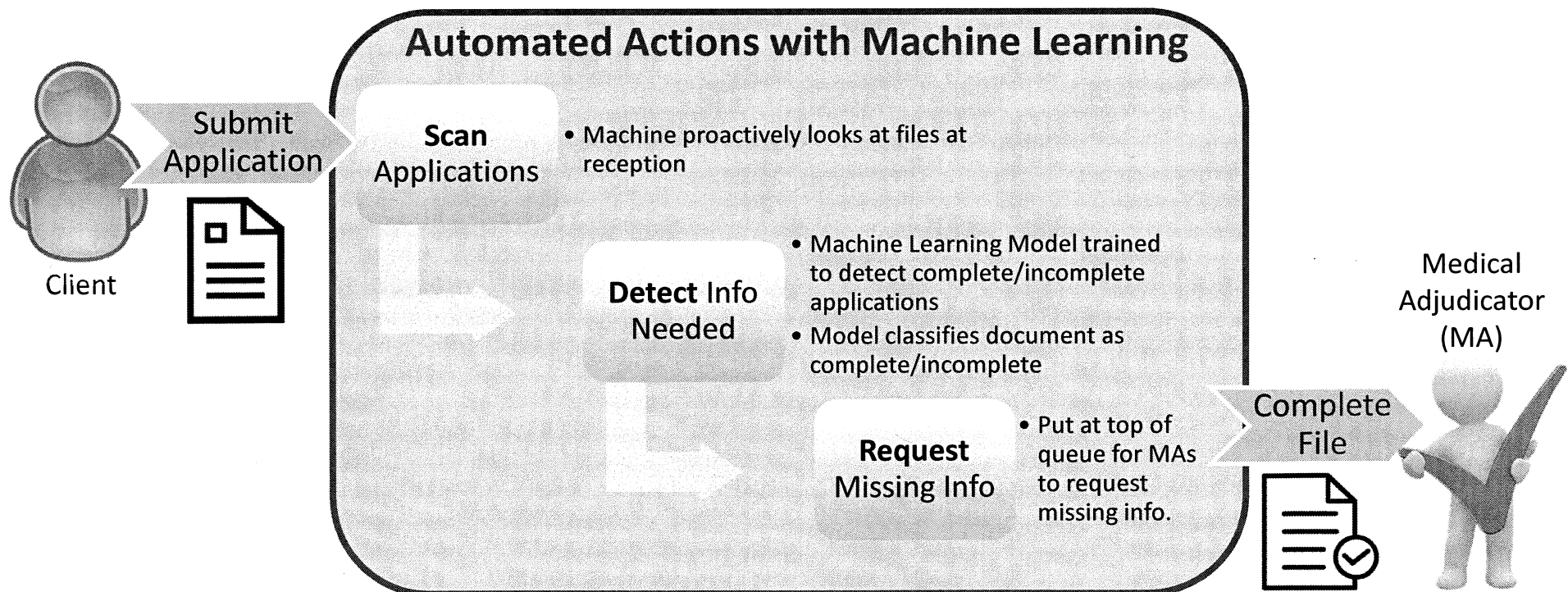
The Canada Pension Plan (CPP) provides disability benefits to people who are disabled and cannot work at any job on a regular basis through the CPPD program. Medical adjudicators (MA) determine eligibility after reviewing detailed applications involving significant medical documentation

### Problem

Applicants have to wait for a MA review to determine if they require additional documentation. Often the MA will not look at an application for several months, making it difficult to meet the service standard of 120 days when extra information is needed.

### Value

- Clients get their benefits faster
- Fewer rejections and appeals due to missing information
- MA's time more efficiently used to evaluate complete files



# Automating Manual Processes

13  
Business  
Context

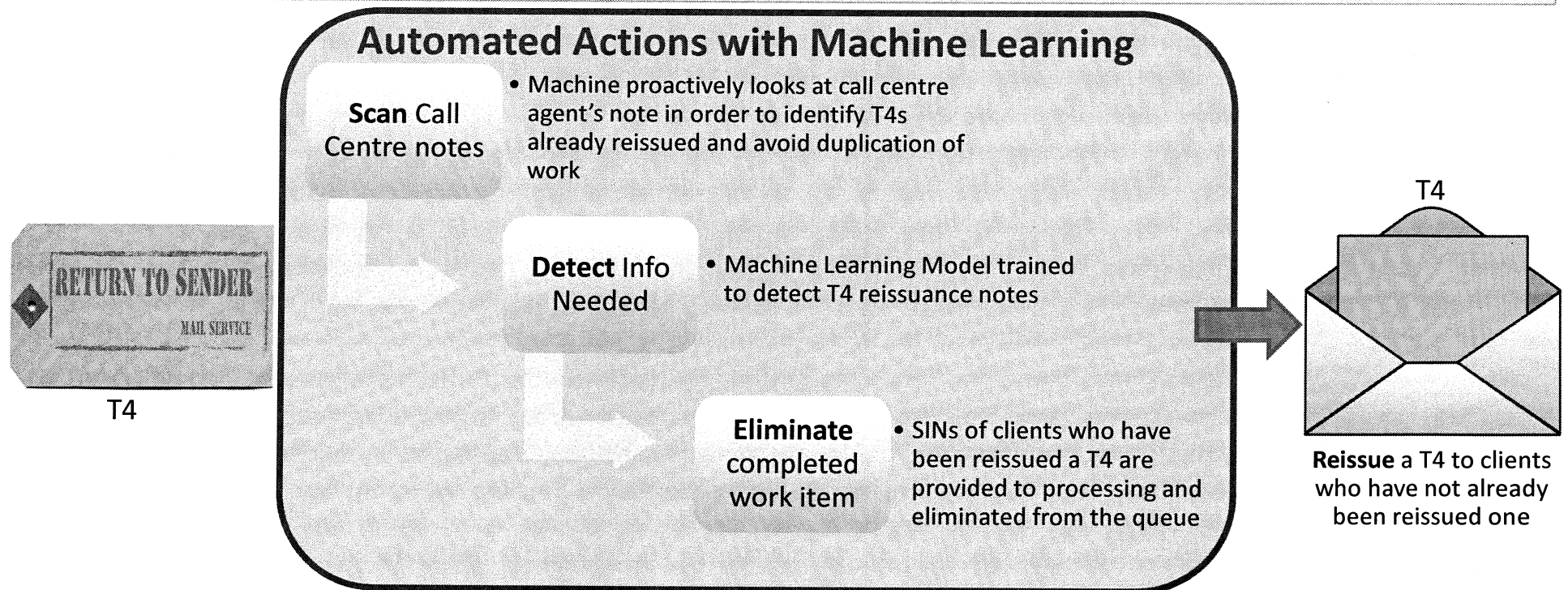
Each year Canada Pension Plan (CPP) and Old Age Security (OAS) recipients receive a T4 form containing the benefits received that they will need to report on their tax return. If they are recipients of the CPP the slip is a T4A(P) while it's a T4A (OAS) for OAS.

Problem

Each year many T4s are returned to the Service Canada processing network due to address changes or other reasons. A large number of clients previously followed-up with Service Canada to request a duplicate T4, which is immediately reissued. Processing still performs resource intensive investigation of each returned item individually.

Value

- Low-value work eliminated: 50,000 work items, or 2.5 FTEs, saved annually, while development only used 0.1 FTE
- Work inventory is reduced, with faster processing for remaining clients
- Once built, the models can be repurposed to answer new questions



14

# Leveraging unstructured data (text, images, audio)

Business  
Context

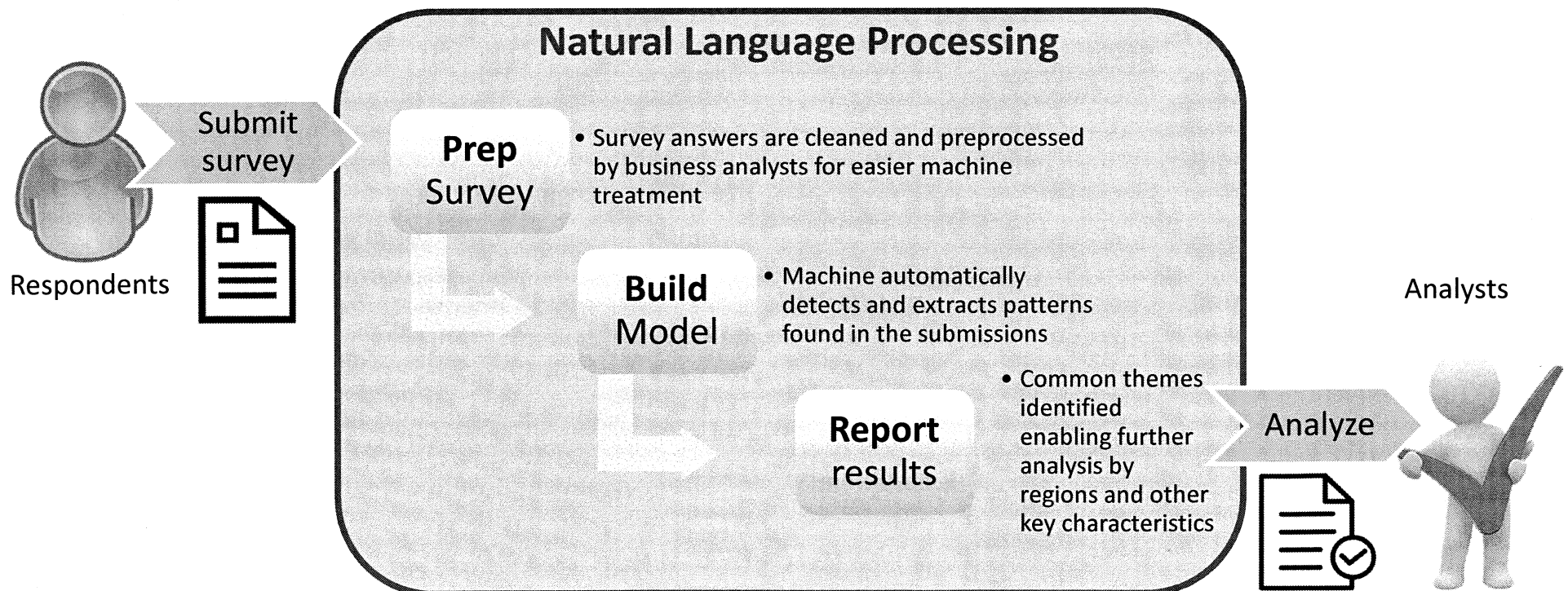
ESDC runs numerous consultations that contain at least one open ended section where participants can express themselves on a specific topic. In 2017, ESDC ran a consultation around the theme of poverty reduction. As part of this consultation, the department received over 800 submissions in free text (emails, online forum and story submissions).

Problem

To get insight into the Poverty Reduction Strategy, analysts would read through different types of submissions (emails, online forums and surveys) from over 800 respondents and try to extract key information manually. This method is resource intensive and inconsistent.

Value

- Fast, unbiased and reproducible insights (took one student 3 weeks to build a model of themes)
- Scalable to very large data sets
- Analysts can delve into explaining results rather than identifying trends saving about 1/6 FTE that would have been allocated to reading the submissions.





15

## Preparing for future technology

### Business Context

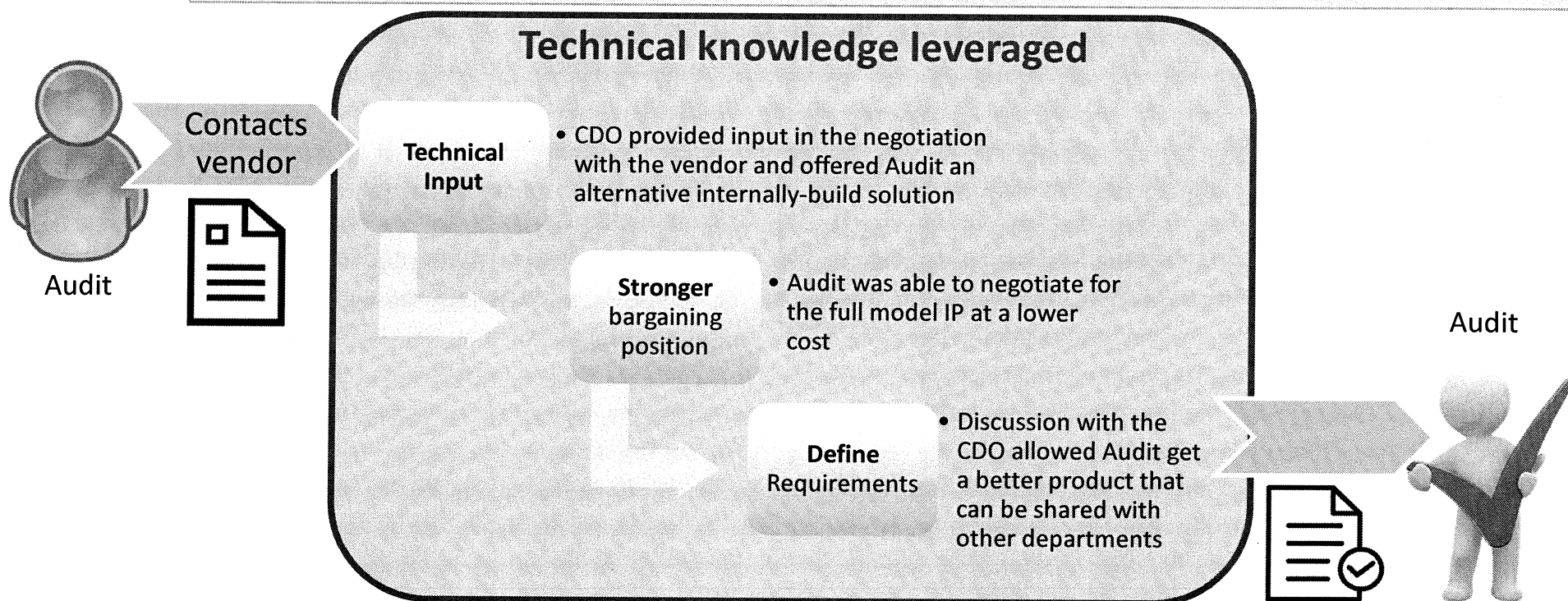
External Artificial Intelligence (AI) tools are increasingly being considered by various branches in their attempt to leverage their data assets. However, the knowledge of the different techniques in the department is sparse so leveraging external vendors is a common option explored.

### Problem

Audit contacted a private vendor to acquire a machine learning categorization tool. The private vendor would have required the department to pay development costs up front, acquire a subscription to use the model and ESDC would own no intellectual property (IP).

### Value

- IP rests with ESDC allowing the model to be repurposed, reused, and shared with other Departments
- Audit contracted for a better product at the same costs
- External expertise is leveraged to hit tight timelines



## Next Steps

- Put models into **production** to deliver value
  - Working closely with the Transformation and Integrated Service Management Branch operations
- Support **Service Transformation Plan** activities
  - Regular participation in the acceleration hub and advising on specific initiatives
- Establish **Analytics Program** throughout 2018
  - Built in collaboration with stakeholders across the organization
- Finalize the **Artificial Intelligence Strategy**
- Work with partners across the organization to discover where data science can help them and **build organizational capacity**
  - Leading data and research streams of 2017-18 post-secondary recruitment to identify strong technical candidates for all of ESDC
- Work with the Innovation, Information and Technology Branch to make **AI sustainable**
  - Developing long-term solution for the models, how they are stored, accessed and maintained



## ANNEX A: Glossary

- **Data science** uses a wide-range of analytic techniques on large amounts of granular data to solve business problems.
- **Advanced Analytics** is very similar to data science but focuses on the techniques rather than the overall problem-solving function.
- **Artificial Intelligence** refers to the ability of computers to complete tasks and make decisions that require human-level judgement. Current AI often makes use of machine learning.
- **Chatbot** refers to an interactive digital question and answer tool.
- **Machine Learning** refers to computer algorithms that are able to learn how to solve specific problems through exposure to data and can improve over time as more data is acquired.
- **Natural language processing** refers to computer algorithms that deal with the intake, interpretation, summarization and discourse of natural language (both written and spoken).
- **Sentiment Analysis** refers to the use of algorithms to identify and extract the emotional reaction of the speaker or writer to an event or a document.



## ANNEX B: How do machines learn?

- The idea behind machine learning is to train machines the way someone would train a human – through experience and expert guidance
- Supervised learning involves an analyst manually coding examples (a training set) that lets the machine develop an algorithm to mirror those decisions
  - For example, assign a number of documents to group 1, 2, or 3 that allows the machine to see a pattern that it uses to classify future documents
- Unsupervised learning involves the machine being given a defined task without human assisted examples of the “right” and “wrong” answers
  - For example, divide these documents into 4 groups without specifying the groups, based on whatever the computer sees as relevant
- Reinforcement learning is where the machine is able to interact with its environment to determine whether or not it achieved its desired outcome and then adjusts its decision making to improve
  - For example, when Netflix recommends a movie and then the user stops watching after 5 minutes, the algorithm learns that that wasn't a good recommendation and refines its algorithm





## Annex C: Description of projects by stakeholders (1 of 8)

Stakeholders	Projects
<b>Human Resources Service Branch</b> <i>(Project Completed)</i>	<p><b>Business Context:</b>            In order to improve employee retention and identify shortcomings in employer-employee relations, HRSB runs an Employee Exit Survey to all employees leaving the department. Some of the fields in this survey are free text.</p> <p><b>Current situation:</b>            Extraction of key information would need to be done manually, which is inconsistent and resources intensive.</p> <p><b>Solution Provided:</b>            Automated the analysis of the employee exit survey using topic modelling. Common themes that may have otherwise fallen through the cracks were identified (i.e. the link between mental health and housing). Using natural language processing is faster, unbiased, scalable and reusable.</p>
<b>Internal Audit Service Branch (IASB)</b> <i>(Project Completed)</i>	<p><b>Business Context:</b>            External Artificial Intelligence (AI) tools are increasingly being considered various branches in their attempt to leverage their data assets. However, the knowledge of the different techniques in the department is sparse so leveraging external vendors is a common option explored.</p> <p><b>Current Situation:</b>            Audit contacted a private vendor to acquire a machine learning categorization tool. The private vendor would have required the department to pay development costs up front, acquire a subscription to use the model and ESDC would own no intellectual property (IP).</p> <p><b>Solution Provided:</b>            CDO provided input in the negotiation with the vendor and offered Audit an alternative internally-build solution. Audit was able to negotiate for the full model IP with no ongoing subscriptions fees. The discussion between the CDO and Audit allowed Audit to better define their requirements.</p>

## Annex C: Description of projects by stakeholders (2 of 8)

Stakeholders	Projects
<p><b>Income Security and Social Development Branch (ISSD) (CPPD)</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b> Canada Pension Plan (CPP) provides disability benefits to people who have made enough contributions to the CPP and who are disabled and cannot work at any job on a regular basis. Medical adjudicators (MA) determine, after reviewing your application, your eligibility.</p> <p><b>Current Situation:</b> CPPD applications have to wait for a medical adjudicator's (MA) review to determine if they require additional documentation. Often this review takes place after several months making it hard to meet the service standard of 120 days.</p> <p><b>Solution Provided:</b> Predicting which client files require additional documentation at receipt to reduce delays and support better and more consistent decisions by CPP-D agents to eliminate costly appeals of decisions.</p>
<p><b>Income Security and Social Development Branch (ISSD) (Homelessness Strategy)</b></p> <p><i>(Project Completed)</i></p>	<p><b>Business Context:</b> ESDC runs, year after year, numerous consultations which almost always contain at least one open ended section where participants can express themselves on a specific topic. In 2017, ESDC ran a consultation around the theme Homelessness. As part of this consultation, the department received hundreds of submissions in free text.</p> <p><b>Current Situation:</b> To get insight in the Homelessness Strategy, analysts would read through all submissions and try to extract key information manually. This method is resource intensive, inconsistent and not reproducible.</p> <p><b>Solution Provided:</b> Through topic modelling, the CDO provided with a fast, scalable and efficient way to extract key information.</p>

## Annex C: Description of projects by stakeholders (3 of 8)

Stakeholders	Projects
<p><b>Labour Program – Information Retrieval System</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b> About 1 million employees in Canada work in federally regulated occupations. The Labour Program promotes cooperation and fairness and provides expert advice and assistance on labour relations matters to workplaces within the federal jurisdiction. To that intent, the Labour Program has developed numerous services, measures and initiatives to assist employers and employees in creating and maintaining a workplace that is conducive to good industrial relations.</p> <p><b>Current Situation:</b> The labour program, upon requests on collective agreements related to federal jurisdiction occupations, need to extract information manually from a sample of collective agreements because of the large number of agreements.</p> <p><b>Solution Provided:</b> Creating a question and answer tool to categorize and facilitate responses to queries of collective agreements. Model allows for a faster, more efficient and reproducible way of querying collective agreements. The model required 0.2 FTE to create and can be reused over and over at almost no costs.</p>
<p><b>Legal Services (Justice Canada)</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b> External A.I. tools are more and more considered by different branches in their attempt to leverage their data assets. Legal is facing an ever growing volume of documentation and need to keep up with the private sector. However, the knowledge of the different techniques in the department is sparse.</p> <p><b>Current Situation:</b> Legal is interested in acquiring an A.I. solution to keep up with the ever increasing amount of documentation. Legal lacks the proper knowledge to define their requirements in terms of A.I. to external vendors.</p> <p><b>Solution Proposed:</b> Providing technical expertise to advise on Artificial Intelligence projects with external contractors to facilitate the acquisition of the right solution.</p>



## Annex C: Description of projects by stakeholders (4 of 8)

Stakeholders	Projects
<p><b>Strategic and Service Policy Branch (SSPB) – Poverty Reduction Strategy</b></p> <p><i>(Project Completed)</i></p>	<p><b>Business Context:</b> ESDC runs, year after year, numerous consultations which almost always contain at least one open ended section where participants can express themselves on a specific topic. In 2017, ESDC ran a consultation around the theme Poverty Reduction. As part of this consultation, the department received hundreds of submissions in free text.</p> <p><b>Current Situation:</b> To get insight in the Poverty Reduction Strategy, analysts would read through all submissions and try to extract key information manually. This method is resource intensive, inconsistent and not reproducible.</p> <p><b>Solution Provided:</b> Automated the extraction of key themes from hundreds of online submissions related to the Poverty Reduction Strategy.</p>
<p><b>Service Transformation Plan (STP)</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b> ESDC is in the midst of transforming the way we deliver services to Canadians. Among the transformation envisioned are modern authentication methods and automated dialoguing solutions.</p> <p><b>Solution Envisioned:</b> Exploring speech recognition technology to identify how it can be used for both client identification as well as analytics.</p>
<p><b>Transformation and Integrated Service Management Branch (TISMB) – OAS related mail</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b> ESDC is in the midst of transforming the way we deliver services to Canadians. Among the transformation envisioned are modern authentication methods and automated dialoguing solutions.</p> <p><b>Current Situation:</b> A significant amount of emails relating to OAS are hitting the system. Many of those are related to work that cannot be actioned or duplicate work that is currently underway or has been addressed. The current email triaging requires many hours of manual effort, which could otherwise be used in more important areas of processing.</p> <p><b>Solution Provided:</b> Through a number of techniques, the CDO streamlined, closed and aggregated email workflow by triaging emails as they arrived.</p>

## Annex C : Description of projects by stakeholders (5 of 8)

Stakeholders	Projects
<p><b>Transformation and Integrated Service Management Branch – OAS overpayment</b></p> <p><i>(Project Completed)</i></p>	<p><b>Business Context:</b> Occasionally under/overpayments are issued to Canadians and Service Canada is notified by the clients. The errors are noted in a system. In order to ensure that every Canadians gets what he/she is entitled too, intensive manual efforts are needed.</p> <p><b>Current Situation:</b> In a number of cases, overpayments and underpayments are issued by ESDC to clients. Information Technology Renewal Delivery System (ITRDS) Notes contains information on those over/underpayments. The Transformation and Integrated Service Management Branch employees are manually going through files and extract the information from the ITRDS Notes. This work is time consuming, resource intensive and inconsistent.</p> <p><b>Solution Provided:</b> The CDO used predictive models to flag ITRDS Notes reporting an over/underpayment amount over a certain threshold. The model demonstrated accurately that it could extract specific information. The model build by the CDO using only 1/3 FTE is efficient, consistent, is replicable and frees up valuable analysts time. Also, the model developed here can be repurposed to do similar task internally at ESDC or in other departments facing similar issues.</p>
<p><b>Transformation and Integrated Service Management Branch – T4 Processing</b></p> <p><i>(Project Completed)</i></p>	<p><b>Business Context:</b> Each year Canada Pension Plan (CPP) and Old Age Security (OAS) recipients receive an information slip containing the information they will need to report on their tax return. If they are recipients of the CPP the slip is a T4A(P) while it's a T4A (OAS) for OAS.</p> <p><b>Current Situation:</b> Service Canada processing network receives numerous returned T4's due to changes in client's address or for other reasons. A significant number of clients follow-up with Service Canada to request a duplicate tax slip. Processing is not made aware which in return create intensive manual investigation.</p> <p><b>Solution Provided:</b> The model identifies when a client's T4 as being reissued, provided the SIN associated with that T4 to processing and eliminates it from the queue. 50,000 work items, or 2.5 FTEs, saved annually, while development only used 0.1 FTE. Work inventory is reduced, with faster processing for remaining clients. Once built, the models can be repurposed to answer new questions.</p>

## Annex C : Description of projects by stakeholders (6 of 8)

Stakeholders	Projects
<p><b>Transformation and Integrated Service Management Branch – EI Application</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b> ESDC is still relying heavily on paper application for the numerous programs in its portfolio. Among them, the Employment Insurance (EI) program requires Canadians to fill a paper application.</p> <p><b>Current Situation:</b> EI application are manually scanned and tags are manually assigned to categorized them.</p> <p><b>Solution Proposed:</b> Automating the classification of EI applications allowing the reallocation of valuable employee time.</p>
<p><b>Transformation and Integrated Service Management Branch – Involuntary Separation – GIS</b></p> <p><i>(Project Completed)</i></p>	<p><b>Business Context:</b> The GIS provides a monthly non-taxable benefit to Old Age Security (OAS) pension recipients who have low income and are living in Canada. In January 2017, guidelines for Guaranteed Income Supplement (GIS) were changed. The GIS guidelines, in the case of a partner admitted to long-term care facility would no longer allow those affected to be assessed as 'single'. This situation leaves one person to pay both household costs and long-term care facility expenses with less GIS support. The Minister of Families, Children and Social Development announced that the guidelines were to be changed to what they were prior to January 2017.</p> <p><b>Current Situation:</b> After a policy change, a number of applicants were denied GIS. The change in policy was reversed and analysts would need to go through all ITRDS notes to identify the clients that were denied.</p> <p><b>Solution Provided:</b> The CDO built a model to identify those automatically, saving numerous FTEs that can be allocated to other tasks.</p>
<p><b>Innovation, Information and Technology Branch (IITB), Shared Service Canada (SSC) (Data Lake)</b></p> <p><i>(Project Underway)</i></p>	<p><b>Current Situation:</b> ESDC analysts and external researchers requiring protected data must go through a lengthy and complicated process to access data.</p> <p><b>Solution Proposed:</b> Exploring a Protected B Cloud Pilot on a variety of data sources including image and speech to provide easier access and advanced analytical tools to researchers.</p>



# Annex C : Description of projects by stakeholders (7 of 8)

Stakeholders	Projects
<p><b>Public Affairs and Stakeholder Relations Branch (PASRB) – Media Monitoring</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b>            Everyday, thousands of articles and news stories on ESDC or its Ministers are produced. Keeping track of what is being said about the department, its programs or its Ministers with manual techniques is becoming more and more difficult as the speed at which news are produced.</p> <p><b>Current Situation:</b>            The Government of Canada (GoC), through Public Safety Canada, purchases a list of articles pertaining to all federal departments. Those lists are then shared with departments for further analysis. ESDC employees currently come in to work at 3 A.M. in order to open and read between 800 and 1000 articles. Of those, about 100 are evaluated and summarized for ESDC’s higher management. Manual monitoring of a sample of all news related to the department is resource intensive and time consuming.</p> <p><b>Solution Proposed:</b>            The CDO is working on a tool that will rate not only a sample, but all the articles provided by Public Safety Canada everyday. It will rate articles according to their content, the mention of an ESDC program or if ESDC’s Ministers are mentioned. Once the monitoring part of the model is up and running, an automatic summarizer will be implemented.</p>

## Annex C : Description of projects by stakeholders (8 of 8)

Stakeholders	Projects
<p><b>Citizen Service Branch(CSB) – MSCA/MSCBA Operations team</b></p> <p><i>(Project Completed)</i></p>	<p><b>Business Context:</b>          My Service Canada Account (MSCA) is an online portal administered by Service Canada, providing convenient and secure access to information regarding: EI, CPP, OAS          MSCA has an online feedback form where users can submit suggestions on how to improve MSCA by filling out this form. The feedback form comments provide valuable information to help improve the MSCA services.</p> <p><b>Current Situation:</b>          MSCA analysts spend up to 25 hours on a monthly basis manually analyzing client feedback. Categorization and analysis is an important task, but the manual process takes away time from MSCA team analysts conducting more in-depth work.</p> <p><b>Solution Proposed:</b>          The research project conducted by the Service Research Division (SRD) was successful in applying machine learning techniques to MSCA feedback form comments in order to facilitate automation of comment categorization. The process of user-feedback categorization now takes a few minutes instead of hours. The time- saving results in more resources being devoted to addressing user concerns and improving government services. SRD also developed an intuitive graphical user interface to interact with the model and is currently investigating visual and report-generating tools to better analyze MSCA user feedback.</p>


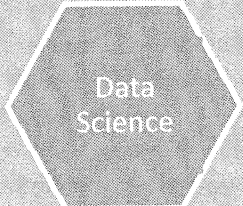
## Annex D : Objective of a Data Strategy

- The main objective of the data strategy is to get data into the hands of people who can **drive value** with the work that they do.
- There are 6 work streams that will make that happen in a secure way, that respects the privacy of individuals, **proving that data can be both more secure and more accessible**.
- Two work streams in particular Data Access and Data Science, will enable ESDC employees and partners such as members of ESDC the Canadian Research Data Centre Network, to perform analytics and research that **will drive both our policy and service mandates**.





## Annex E: Status: Progress on leveraging our data assets...

 <p>Access</p>	<ul style="list-style-type: none"> <li>Enhancing collaboration with <b>Statistics Canada, academia, and other departments:</b> <ul style="list-style-type: none"> <li>Strengthened joint ESDC-StatCan governance (e.g., ADM-ACS, DM-CS)</li> <li>ESDC-StatCan Data Sharing Roadmap approved and implementation underway</li> <li>Created Interdept. Working Group on Children's Data to increase data sharing and address gaps</li> </ul> </li> <li>Public release of over 100 data sets as <b>open data</b></li> <li>Development and early implementation of a <b>Hackathon Strategy</b> (e.g., on-track for a poverty reduction hackathon Spring 2018)</li> <li>Working with Justice Canada and other stakeholders to identify barriers and provide recommendations to enable data sharing through <b>Privacy Act Reform</b></li> <li>Chief Data Officer co-chairing new <b>Privacy and Data Access Committee</b></li> </ul>
 <p>Data Science</p>	<ul style="list-style-type: none"> <li>Building and piloting new approaches to leveraging our data           <ul style="list-style-type: none"> <li><b>Leveraging unstructured data</b> (e.g., eliminated 50,000 pieces of work previously done manually by revamping the returned T4 slips process, automated email triage of Old Age Security clients, thematic analysis of online Poverty Reduction Strategy Consultation)</li> <li><b>Supporting proactive decisions</b> (e.g., Predicting which client files require additional documentation at receipt to reduce delays and support better decision-making by CPP-D agents to eliminate costly appeals of decisions)</li> <li><b>Preparing for future technologies</b> (e.g., developing Analytics Program and Artificial Intelligence strategy, exploring Speech Recognition technology to identify how it can be used for client identification and analytics)</li> </ul> </li> </ul>

See Annex A for detailed overview of progress milestones



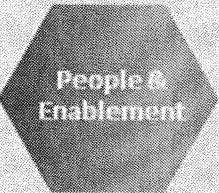


Employment and  
Social Development Canada

Emploi et  
Développement social Canada

Canada

## ... while building key data foundations

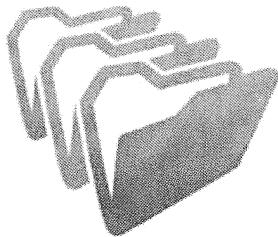
	<ul style="list-style-type: none"> <li>• Piloting formalized <b>data stewardship</b> function with HR to demonstrate the value of a systematic process to identify and resolve data issues.</li> <li>• Piloting improved tools and methodology for assessing <b>data quality</b> to improve quality of performance information reporting.</li> <li>• Develop performance metrics proposal to articulate the anticipated outcomes from enhanced data governance.</li> </ul>
	<ul style="list-style-type: none"> <li>• <b>Enterprise Business Information and Analytics Reference Architecture</b> proposal developed with Innovation, Information and Technology Branch (IITB) and approved by the ESDC Enterprise Architecture Review Board (EARB)</li> <li>• Chief Data Office leading Service Transformation Plan <b>Single Client View “pod”</b></li> </ul>
	<ul style="list-style-type: none"> <li>• <b>Post-Secondary Recruitment</b> campaign for the data and research streams likely to result in a much larger pool of high quality candidates at the masters/PhD level with 120 advancing to the interview phase (compared to a pool of 15 in 2017).</li> <li>• <b>Collaborating with GoC stakeholders</b> to advance data initiatives (e.g. DG Data Leads Committee, Metadata Standards Sub-group, Open Government Multi-Stakholder Forum, hosting 2018 GoC Data Conference)</li> <li>• Developing <b>Communications &amp; Engagement and Data Literacy</b> plans</li> </ul>

See Annex A for detailed overview of progress milestones



## Annex F : To What Purpose is ESDC using Data Science?

### Categorization



Dividing the data into groups, categories, or topics that are related

- Triaging work
- Segmenting clients
- Identifying topics or themes

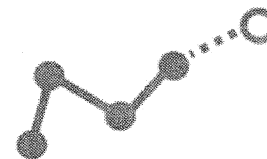
### Key Information Extraction



Isolating important information and extracting the useful part

- Chatbots\*
- Autofill forms or templates
- Descriptive information

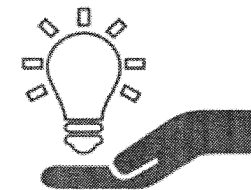
### Prediction



Using historical relationships to predict future outcomes

- Advise agents on decision to be taken
- Predict future workload
- Predict potential benefits for clients

### Technical Expertise to Advise



Exploring emerging technologies and approaches to inform decisions

- Help ESDC sign better contracts
- Identify key emerging tech and adapt it for ESDC's needs.
- Repurpose and adapt models for quick wins

\* See glossary in annex A



It's time to do more <with data>

# Data Science and A.I. at ESDC

Chief Data Office, ESDC  
Last updated: October 15, 2018



Employment and  
Social Development Canada

Emploi et  
Développement social Canada

Canada

# Outline

1. Context
2. ESDC Data Strategy
3. Defining A.I.
4. A.I. Pilot Projects
5. ESDC A.I. Strategy
6. Going forward



# ESDC is adapting to a changing data-driven world

- Client expectations have changed

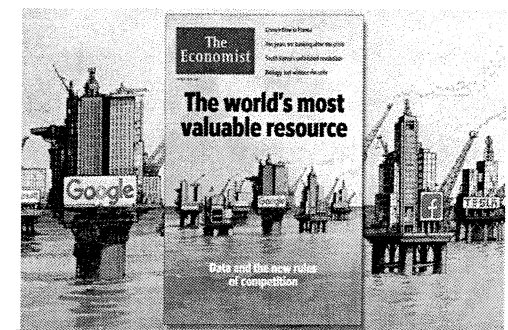
Clients expect private sector level of service, but with a much higher standard for privacy protection. We must integrate our data to support proactive client service within secure and managed environments.

- ESDC has changed

Evidence-based decision making and Results & Delivery, the Service Transformation mandate, and a focus on transparency require access to data. We must know what data we have and how to use it.

- Technology and analytical methods have changed

We have fallen behind in the underlying investments needed to use and extract the value of data. We need people, technology and an analytics program to tie it all together.



**"I believe that government departments and organizations urgently need to turn their attention to this issue. They need to focus on collecting the right data to support their activities, on ensuring that data is well-managed and up-to-date, and on fully using this data not only to inform their core business, but also to support reporting and continuous improvement."**

*2016 Spring Reports of the Auditor General of Canada, Opening Statement, May 3, 2016*



Employment and  
Social Development Canada

Emploi et  
Développement social Canada



# Our definition of Artificial Intelligence (A.I.)

- A.I. solutions are: digital solutions that exhibit human or higher-level judgement to carry out tasks
- Must fall into one or more of the following modern A.I. domains
  - Natural Language Processing (NLP)
  - Computer vision
  - Audio processing
  - Cognitive modeling
  - Strategic optimization
- At some level, must contain machine learning elements that enable them to continually improve their ability to carry out their task



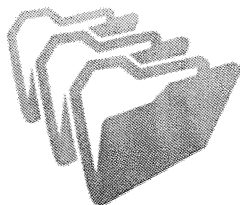
# How do machines learn?

- The idea behind machine learning is to train machines the way someone would train a human – through experience and expert guidance. There are three main learning approaches:
- **Supervised learning** involves an analyst manually coding examples (a training set) that lets the machine develop an algorithm to mirror those decisions
- **Unsupervised learning** involves the machine being given a defined task without human assisted examples of the “right” and “wrong” answers
- **Reinforcement learning** is where the machine is able to interact with its environment to determine whether or not it achieved its desired outcome and then adjusts its decision making to improve



# The need for A.I. expertise at ESDC

## Categorization



Dividing the data into groups, categories, or topics that are related

- Triaging work
- Segmenting clients
- Identifying topics or themes

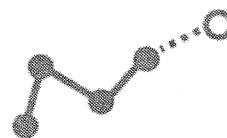
## Key Information Extraction



Isolating important information and extracting the useful part

- Chatbots\*
- Autofill forms or templates
- Descriptive information

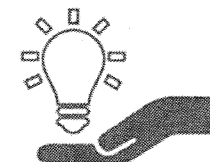
## Prediction



Using historical relationships to predict future outcomes

- Advise agents on decision to be taken
- Predict future workload
- Predict potential benefits for clients

## Technical Expertise to Advise



Exploring emerging technologies and approaches to inform decisions

- Help ESDC sign better contracts
- Identify key emerging tech and adapt it for ESDC's needs.
- Repurpose and adapt models for quick wins

\*Chatbot: or "bot" is an application that performs an automated task, like interacting and responding to users using natural language processing, and performing automated tasks



Employment and  
Social Development Canada

Emploi et  
Développement social Canada

# CDO Partnered A.I. Pilot Projects

- HR
  - Employee exit survey analysis (*Completed*)
  - Onboarding chatbot (*Ongoing*)
  - Application screening (*Ongoing*)
- Operations
  - T4 returns automation (*Completed*)
  - Workload triage (*Ongoing*)
  - Agent case notes NLP\* tool (*Ongoing*)
  - Various processing pilots (*Ongoing*)
  - GIS Involuntary Separation (*Completed*)
- Audit
  - Horizontal risk assessment (*Phase 1 Completed*)
- Communications
  - Newsdesk automation (*Ongoing*)
  - Stakeholder monitoring tool (*Not Started*)
- Labour program
  - Collective agreement information retrieval (*Completed*)
- Legal services
  - Paralegal support for legal files (*Ongoing*)
- Survey Analysis
  - Poverty reduction strategy (*Completed*)
  - Homelessness (*Completed*)

\*NLP: Natural Language Processing: refers to the ability of machines to read, understand, categorize, summarize, extract information from and create information in written natural language.



# Example #1: A.I.-enabled work item processing

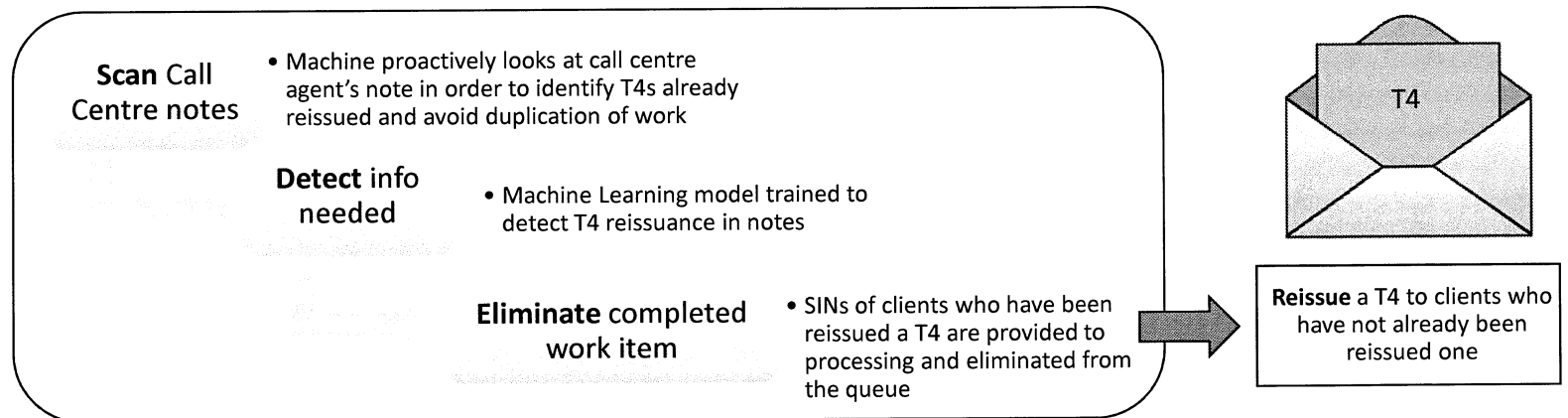
## Business Context

Canada Pension Plan (CPP) and Old Age Security (OAS) recipients receive annually a T4 form containing the benefits received that they will need to report on their tax return. Many T4s are returned to Service Canada due to address changes or other reasons and must be investigated.

## Problem

Each year, large numbers of clients follow up with Service Canada to request a duplicate T4, which is immediately reissued. However, the reissuance is only captured in call centre notes and the resource intensive investigation of every single individual returned T4 is still performed.

## Solution



## Value

- Low-value work eliminated: 50,000 work items, or 2.5 FTEs, saved annually, while development only used 0.1 FTE
- Work inventory is reduced, with faster processing for remaining clients
- Once built, the models can be repurposed to answer new questions



Employment and  
Social Development Canada

Emploi et  
Développement social Canada

# Example #2: A.I.-enabled media monitoring tool

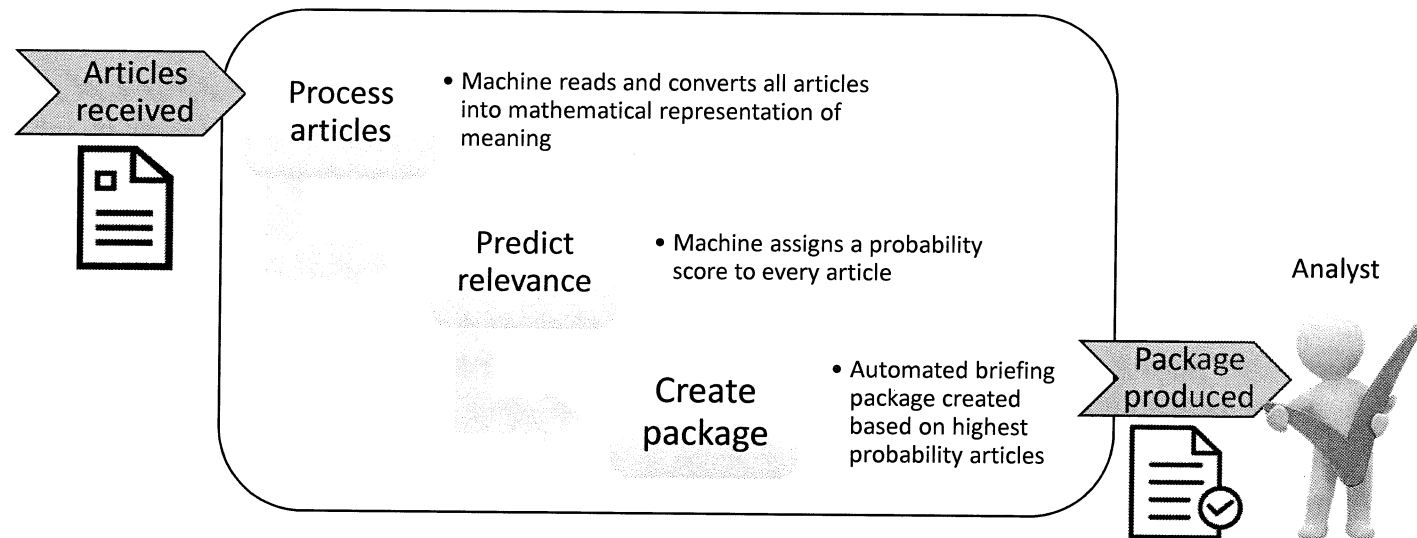
## Business Context

Every day, a package is produced for the ministers that provides pertinent news articles pertaining to ESDC's mandate.

## Problem

The package is produced by analysts who manually review news media starting at 3am each day. The volume to be monitored is often beyond the capacity to review every article and leaves little time for analysis.

## Solution



## Value

- Reduced workload for employees
- Timelier package production allows for real analysis on what the news means for the Department
- More media can be searched and evaluated in a consistent fashion



Employment and  
Social Development Canada

Emploi et  
Développement social Canada



# Example #3: A.I. vendor negotiations with Audit

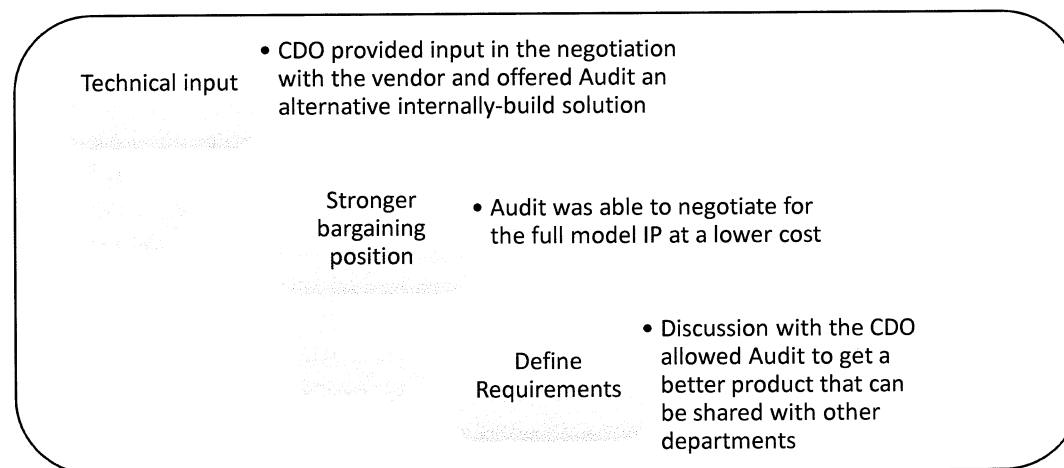
## Business Context

External Artificial Intelligence (A.I.) tools are increasingly being considered by various branches in their attempt to leverage their data assets. However, expertise in A.I. is still sparse in the department, making negotiations with external vendors challenging.

## Problem

Audit contacted a private vendor to acquire a machine learning categorization tool. The private vendor would have required the department to pay development costs up front, acquire a subscription to use the model and ESDC would own no intellectual property (IP).

## Solution



## Value

- IP rests with ESDC allowing the model to be repurposed, reused, and shared with other Departments
- Audit contracted for a better product at the same costs
- External expertise is leveraged to hit tight timelines



Employment and  
Social Development Canada

Emploi et  
Développement social Canada

# Lessons Learned from the pilot projects – Partnership Matters

- Joint team of business and technology experts
  - Business needs to invest has much time in the development as the technology experts.
- Data team working with Business units to solve problems
  - The pilot projects undertaken by the CDO with its partners were to solve real business problems, its not innovation for the sake of innovation.
- Agile workflow - Enabling feedback in the development process
  - The data scientists and business experts work hand in hand to advance projects as a coordinated team.
- The need for internal technical experts
  - Internal capacity is needed to understand what is feasible and to negotiate with vendors in order to make the most of taxpayer money, understand external solutions and address legal, ethical considerations.
  - Dependency on data is more significant than availability.
- Buy vs. Build
  - Its not one or the other – IT'S BOTH !
  - Often solutions require extensive customization from business input that vendors are not often well positioned to provide as they are not mature in this space

# Partnering Across GoC

- Internally developed A.I. solutions could be repurposed to deliver value to other departments.
- No processes currently exist to scale internal open source development GoC wide.
- The CDO is exploring various collaboration and funding approaches.
- Three projects have attracted significant interest from other departments:
  - Risk Insight tool (Audit) *(Details on the project on slide 16)*
  - Media monitoring tool (PASRB) *(Details on the project on slide 22)*
  - HR application screening tool (HRSB) *(Details on the project on slide 23)*

# ESDC's A.I. Strategy

- The need for an A.I. Strategy:
  - Clarify thinking around challenges and opportunities
  - Support GoC-wide A.I. policies
  - Educate and dispel myths
  - Help ensure responsible use of A.I.
- Grounded in experience
- Describes domains of A.I. investment and applications in ESDC
- A.I. Policy addresses key aspects, including:
  - Legal, Ethics, Transparency, Accountability, Privacy, and Security



## ESDC's A.I. Strategy Principles

1. Develop a modern AI suite to transform the way ESDC delivers services to Canadians
2. Develop a policy for acceptable AI use in light of the risks it poses
3. Strengthen our internal capacity in AI development
4. Organize ourselves to properly steward the most important component of the current AI wave: the data
5. Obtain maximum public value for our investments in AI in discussions with vendors



## ESDC's A.I. Strategy Principles

6. Engage across the organization to promote AI and coordinate initiatives
7. Put in place the right platform for the development and deployment of AI solutions
8. Develop processes and controls for AI models to ensure they do what we want them to do
9. Design a framework for monitoring performance and evaluating success of AI solutions to prove value to Canadians





# Going forward

- Finalize the **Artificial Intelligence Strategy**
  - Version 1.0 of the A.I. Strategy is currently being circulated for comments
- Work with partners across the organization to **build organizational capacity**
- Operationalize models in **production environments**
- **Partner** with other GoC organizations
  - Work with external partners to develop and deploy models that can impact many GoC organizations, train and educate employees and provide advices
  - Leverage other department's proposal to fund the CDO's work to augment the CDO's capacity
- Support **Service Transformation Plan** activities
  - Regular participation in the acceleration hub and advising on specific initiatives
- Establish **Analytics Program**



Stakeholders	Projects
<p><b>Internal Audit Service Branch (IASB)</b></p> <p><i>(Project Completed)</i></p>	<p><b>Business Context:</b>            External Artificial Intelligence (AI) tools are increasingly being considered by various branches in their attempt to leverage their data assets. However, the knowledge of the different techniques in the department is sparse so leveraging external vendors is a common option explored.</p> <p><b>Current Situation:</b>            Audit contacted a private vendor to acquire a machine learning categorization tool. The private vendor would have required the department to pay development costs up front, acquire a subscription to use the model and ESDC would own no intellectual property (IP).</p> <p><b>Solution Provided:</b>            CDO provided input in the negotiation with the vendor and offered Audit an alternative internally-build solution. Audit was able to negotiate for the full model IP with no ongoing subscriptions fees. The discussion between the CDO and Audit allowed Audit to better define their requirements.</p>

Stakeholders	Projects
<p><b>Income Security and Social Development Branch (ISSD) (CPPD)</b></p> <p><i>(Project Cancelled)</i></p>	<p><b>Business Context:</b>            Canada Pension Plan (CPP) provides disability benefits to people who have made enough contributions to the CPP and who are disabled and cannot work at any job on a regular basis. Medical adjudicators (MA) determine, after reviewing your application, your eligibility.</p> <p><b>Current Situation:</b>            CPPD applications have to wait for a medical adjudicator's (MA) review to determine if they require additional documentation. Often this review takes place after several months making it hard to meet the service standard of 120 days.</p> <p><b>Solution Provided:</b>            Predicting which client files require additional documentation at receipt to reduce delays and support better and more consistent decisions by CPP-D agents to eliminate costly appeals of decisions.</p>
<p><b>Income Security and Social Development Branch (ISSD) (Homelessness Strategy)</b></p> <p><i>(Project Completed)</i></p>	<p><b>Business Context:</b>            ESDC runs, year after year, numerous consultations which almost always contain at least one open ended section where participants can express themselves on a specific topic. In 2017, ESDC ran a consultation around the theme Homelessness. As part of this consultation, the department received hundreds of submissions in free text.</p> <p><b>Current Situation:</b>            To get insight in the Homelessness Strategy, analysts would read through all submissions and try to extract key information manually. This method is resource intensive, inconsistent and not reproducible.</p> <p><b>Solution Provided:</b>            Through topic modelling, the CDO provided with a fast, scalable and efficient way to extract key information.</p>

Stakeholders	Projects
<p><b>Labour Program – Information Retrieval System</b></p> <p><i>(Project Completed)</i></p>	<p><b>Business Context:</b> About 1 million employees in Canada work in federally regulated occupations. The Labour Program promotes cooperation and fairness and provides expert advice and assistance on labour relations matters to workplaces within the federal jurisdiction. To that intent, the Labour Program has developed numerous services, measures and initiatives to assist employers and employees in creating and maintaining a workplace that is conducive to good industrial relations.</p> <p><b>Current Situation:</b> The labour program, upon requests on collective agreements related to federal jurisdiction occupations, need to extract information manually from a sample of collective agreements because of the large number of agreements.</p> <p><b>Solution Provided:</b> Creating a question and answer tool to categorize and facilitate responses to queries of collective agreements. Model allows for a faster, more efficient and reproducible way of querying collective agreements. The model required 0.2 FTE to create and can be reused over and over at almost no costs.</p>
<p><b>Legal Services (Justice Canada)</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b> External A.I. tools are more and more considered by different branches in their attempt to leverage their data assets. Legal is facing an ever growing volume of documentation and need to keep up with the private sector. However, the knowledge of the different techniques in the department is sparse.</p> <p><b>Current Situation:</b> Legal is interested in acquiring an A.I. solution to keep up with the ever increasing amount of documentation. Legal lacks the proper knowledge to define their requirements in terms of A.I. to external vendors.</p> <p><b>Solution Proposed:</b> Providing technical expertise to advise on Artificial Intelligence projects with external contractors to facilitate the acquisition of the right solution.</p>



Stakeholders	Projects
<p><b>Strategic and Service Policy Branch (SSPB) – Poverty Reduction Strategy</b></p> <p><i>(Project Completed)</i></p>	<p><b>Business Context:</b>            ESDC runs, year after year, numerous consultations which almost always contain at least one open ended section where participants can express themselves on a specific topic. In 2017, ESDC ran a consultation around the theme Poverty Reduction. As part of this consultation, the department received hundreds of submissions in free text.</p> <p><b>Current Situation:</b>            To get insight in the Poverty Reduction Strategy, analysts would read through all submissions and try to extract key information manually. This method is resource intensive, inconsistent and not reproducible.</p> <p><b>Solution Provided:</b>            Automated the extraction of key themes from hundreds of online submissions related to the Poverty Reduction Strategy.</p>
<p><b>Service Transformation Plan (STP)</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b>            ESDC is in the midst of transforming the way we deliver services to Canadians. Among the transformation envisioned are modern authentication methods and automated dialoguing solutions.</p> <p><b>Solution Envisioned:</b>            Exploring speech recognition technology to identify how it can be used for both client identification as well as analytics.</p>
<p><b>Transformation and Integrated Service Management Branch (TISMB) – OAS related mail)</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b>            ESDC is in the midst of transforming the way we deliver services to Canadians. Among the transformation envisioned are modern authentication methods and automated dialoguing solutions.</p> <p><b>Current Situation:</b>            A significant amount of emails relating to OAS are hitting the system. Many of those are related to work that cannot be actioned or duplicate work that is currently underway or has been addressed. The current email triaging requires many hours of manual effort, which could otherwise be used in more important areas of processing.</p> <p><b>Solution Provided:</b>            Through a number of techniques, the CDO streamlined, closed and aggregated email workflow by triaging emails as they arrived.</p>

Stakeholders	Projects
<p><b>Transformation and Integrated Service Management Branch – OAS overpayment</b></p> <p><i>(Project Completed)</i></p>	<p><b>Business Context:</b> Occasionally under/overpayments are issued to Canadians and Service Canada is notified by the clients. The errors are noted in a system. In order to ensure that every Canadians gets what he/she is entitled too, intensive manual efforts are needed.</p> <p><b>Current Situation:</b> In a number of cases, overpayments and underpayments are issued by ESDC to clients. Information Technology Renewal Delivery System (ITRDS) notes contains information on those over/underpayments. The Transformation and Integrated Service Management Branch employees are manually going through files and extract the information from the ITRDS Notes. This work is time consuming, resource intensive and inconsistent.</p> <p><b>Solution Provided:</b> The CDO used predictive models to flag ITRDS notes reporting an over/underpayment amount over a certain threshold. The model demonstrated accurately that it could extract specific information. The model build by the CDO using only 1/3 FTE is efficient, consistent, is replicable and frees up valuable analysts time. Also, the model developed here can be repurposed to do similar task internally at ESDC or in other departments facing similar issues.</p>
<p><b>Transformation and Integrated Service Management Branch – T4 Processing</b></p> <p><i>(Project Completed)</i></p>	<p><b>Business Context:</b> Each year Canada Pension Plan (CPP) and Old Age Security (OAS) recipients receive an information slip containing the information they will need to report on their tax return. If they are recipients of the CPP the slip is a T4A(P) while it's a T4A (OAS) for OAS.</p> <p><b>Current Situation:</b> Service Canada processing network receives numerous returned T4's due to changes in client's address or for other reasons. A significant number of clients follow-up with Service Canada to request a duplicate tax slip. Processing is not made aware which in return create intensive manual investigation.</p> <p><b>Solution Provided:</b> The model identifies when a client's T4 as being reissued, provided the SIN associated with that T4 to processing and eliminates it from the queue. 50,000 work items, or 2.5 FTEs, saved annually, while development only used 0.1 FTE. Work inventory is reduced, with faster processing for remaining clients. Once built, the models can be repurposed to answer new questions.</p>



Stakeholders	Projects
<p><b>Transformation and Integrated Service Management Branch – Involuntary Separation – GIS</b></p> <p><i>(Project Completed)</i></p>	<p><b>Business Context:</b> The GIS provides a monthly non-taxable benefit to Old Age Security (OAS) pension recipients who have low income and are living in Canada. In January 2017, guidelines for Guaranteed Income Supplement (GIS) were changed. The GIS guidelines, in the case of a partner admitted to long-term care facility would no longer allow those affected to be assessed as 'single'. This situation leaves one person to pay both household costs and long-term care facility expenses with less GIS support. The Minister of Families, Children and Social Development announced that the guidelines were to be changed to what they were prior to January 2017.</p> <p><b>Current Situation:</b> After a policy change, a number of applicants were denied GIS. The change in policy was reversed and analysts would need to go through all Information Technology Renewal Delivery System (ITRDS) notes to identify the clients that were denied.</p> <p><b>Solution Provided:</b> The CDO built a model to identify those automatically, saving numerous FTEs that can be allocated to other tasks.</p>
<p><b>Innovation, Information and Technology Branch (IITB), Shared Service Canada (SSC) (Data Lake)</b></p> <p><i>(Project Underway)</i></p>	<p><b>Current Situation:</b> ESDC analysts and external researchers requiring protected data must go through a lengthy and complicated process to access data.</p> <p><b>Solution Proposed:</b> Exploring a Protected B Cloud Pilot on a variety of data sources including image and speech to provide easier access and advanced analytical tools to researchers.</p>

Stakeholders	Projects
<p><b>Public Affairs and Stakeholder Relations Branch (PASRB) – Media Monitoring (Christine)</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b>            Everyday, thousands of articles and news stories on ESDC or its Ministers are produced. Keeping track of what is being said about the department, its programs or its Ministers with manual techniques is becoming more and more difficult as the speed at which news are produced.</p> <p><b>Current Situation:</b>            The Government of Canada (GoC), through Public Safety Canada, purchases a list of articles pertaining to all federal departments. Those lists are then shared with departments for further analysis. ESDC employees currently come in to work at 3 A.M. in order to open and read between 800 and 1000 articles. Of those, about 100 are evaluated and summarized for ESDC's higher management. Manual monitoring of a sample of all news related to the department is resource intensive and time consuming.</p> <p><b>Solution Proposed:</b>            The CDO is working on a tool that will rate not only a sample, but all the articles provided by Public Safety Canada everyday. It will rate articles according to their content, the mention of an ESDC program or it ESDC's Ministers are mentioned. Once the monitoring part of the model is up and running, an automatic summarizer will be implemented.</p>

Stakeholders	Projects
<p><b>Human Resources Service Branch – Application Screening Automation</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b> Screening for selection processes is lengthy and onerous. There are thousands of applicants for entry positions such as PM-01 and AS-01 and too few people to process them.</p> <p><b>Current Situation:</b> Entry positions receive upwards to 5,000 applications which are screened manually. The scope of the screening process is such that not enough employees can do them all in a reasonable amount of time so a random sampling approach is often used. This is straining resources, quite costly, and can often miss quality candidates.</p> <p><b>Solution Provided:</b> The CDO is developing a machine learning tool to can classify candidates based on selection criteria. A classifier will be trained on historical data from past selection processes to screen all applicants. A second model using similarity of answers and content from the curriculum vitae can also be used for ranking and support the predictions of the classifier (ensemble approach). The tool would be deployed as an application that processes the PSRS data and presents results of the screening with descriptive information. Significant attention is being given to the legal and ethical aspects of this tool and is being developed with Legal at the table.</p>
<p><b>Human Resources Service Branch</b></p> <p><i>(Project Completed)</i></p>	<p><b>Business Context:</b> In order to improve employee retention and identify shortcomings in employer-employee relations, HRSB runs an Employee Exit Survey to all employees leaving the department. Some of the fields in this survey are free text.</p> <p><b>Current situation:</b> Extraction of key information would need to be done manually, which is inconsistent and resources intensive.</p> <p><b>Solution Provided:</b> Automated the analysis of the employee exit survey using topic modelling. Common themes that may have otherwise fallen through the cracks were identified (i.e. the link between mental health and housing). Using natural language processing is faster, unbiased, scalable and reusable.</p>



Stakeholders	Projects
<p><b>Human Resources Service Branch</b>  <b>– New Employee Onboarding (NEO) Chatbot</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b>  ESDC is large department that onboards many new employees on an ongoing basis, and the process requires many partners (e.g. security, IITBD, HR, manager). Each new employee must obtain several accounts (e.g. for learning, pay, HR), complete mandatory training, learn about the available tools, like National Service Desk, get access to resources needed for their work, and situate themselves in their new work context.</p> <p><b>Current Situation:</b>  Current tools for onboarding include a checklist, email templates, and the creation and implementation of a plan by the manager. This involves manual checks that mandatory training has been completed, requesting certain accounts, and providing information to the employee about resources and guidelines, on top of introduction to work tasks. In many cases the success of onboarding is determined by a manager's ability to identify additional training, information or orientation needed, which creates demands on the manager's time with conflicting priorities and resulting in an inconsistent onboarding approach across the department.</p> <p><b>Solution Provided:</b>  The CDO in partnership with HRSB is developing a ChatBot/Virtual Assistant for New Employee Onboarding (NEO). NEO will guide employees on what accounts they need, where to find them, where to ask for help or get extra information, and track completed training. The deployment of the bot will standardize the onboarding experience, allow for continued support for new employees, create awareness of ESDC activities, learning experiences and services, as well as allow for tracking of tasks and potential for analytics on the onboarding process.</p>

It's time to do more <with data>

# Data Science and A.I. at ESDC

Chief Data Office, ESDC



Employment and  
Social Development Canada

Emploi et  
Développement social Canada

Canada

# Outline

1. Context
2. ESDC Data Strategy
3. Defining A.I.
4. A.I. Pilot Projects
5. ESDC A.I. Strategy
6. Going forward





# ESDC is adapting to a changing data-driven world

- **Client expectations have changed**

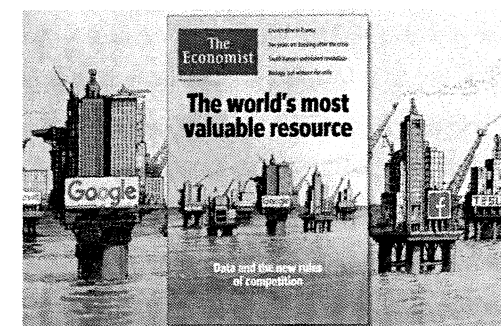
Clients expect private sector level of service, but with a much higher standard for privacy protection. We must integrate our data to support proactive client service within secure and managed environments.

- **ESDC has changed**

Evidence-based decision making and Results & Delivery, the Service Transformation mandate, and a focus on transparency require access to data. We must know what data we have and how to use it.

- **Technology and analytical methods have changed**

We have fallen behind in the underlying investments needed to use and extract the value of data. We need people, technology and an analytics program to tie it all together.



**"I believe that government departments and organizations urgently need to turn their attention to this issue. They need to focus on collecting the right data to support their activities, on ensuring that data is well-managed and up-to-date, and on fully using this data not only to inform their core business, but also to support reporting and continuous improvement."**

*2016 Spring Reports of the Auditor General of Canada, Opening Statement, May 3, 2016*



Employment and  
Social Development Canada

Emploi et  
Développement social Canada

# Our definition of Artificial Intelligence (A.I.)

- A.I. solutions are: digital solutions that exhibit human or higher-level judgement to carry out tasks
- Must fall into one or more of the following modern A.I. domains
  - Natural Language Processing (NLP)
  - Computer vision
  - Audio processing
  - Cognitive modeling
  - Strategic optimization
- At some level, must contain machine learning elements that enable them to continually improve their ability to carry out their task



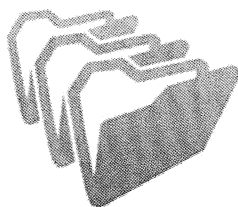
# How do machines learn?

- The idea behind machine learning is to train machines the way someone would train a human – through experience and expert guidance. There are three main learning approaches:
- **Supervised learning** involves an analyst manually coding examples (a training set) that lets the machine develop an algorithm to mirror those decisions
- **Unsupervised learning** involves the machine being given a defined task without human assisted examples of the “right” and “wrong” answers
- **Reinforcement learning** is where the machine is able to interact with its environment to determine whether or not it achieved its desired outcome and then adjusts its decision making to improve



# The need for A.I. expertise at ESDC

## Categorization



Dividing the data into groups, categories, or topics that are related

- Triageing work
- Segmenting clients
- Identifying topics or themes

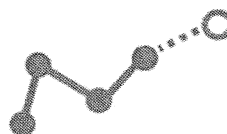
## Key Information Extraction



Isolating important information and extracting the useful part

- Chatbots\*
- Autofill forms or templates
- Descriptive information

## Prediction



Using historical relationships to predict future outcomes

- Advise agents on decision to be taken
- Predict future workload
- Predict potential benefits for clients

## Technical Expertise to Advise



Exploring emerging technologies and approaches to inform decisions

- Help ESDC sign better contracts
- Identify key emerging tech and adapt it for ESDC's needs.
- Repurpose and adapt models for quick wins

\*Chatbot: or "bot" is an application that performs an automated task, like interacting and responding to users using natural language processing, and performing automated tasks



Employment and  
Social Development Canada

Emploi et  
Développement social Canada

# CDO-Partnered A.I. Pilot Projects

- HR
  - Employee exit survey analysis (*Completed*)
  - Onboarding chatbot (*Ongoing*)
  - Application screening (*Ongoing*)
- Operations
  - T4 returns automation (*Completed*)
  - Workload triage (*Ongoing*)
  - Agent case notes NLP\* tool (*Ongoing*)
  - Various processing pilots (*Ongoing*)
  - GIS Involuntary Separation (*Completed*)
- Communications
  - Newsdesk automation (*Ongoing*)
  - Stakeholder monitoring tool (*Not Started*)
- Labour program
  - Collective agreement information retrieval (*Completed*)
- Legal services
  - Paralegal support for legal files (*Ongoing*)
- Survey Analysis
  - Poverty reduction strategy (*Completed*)
  - Homelessness (*Completed*)

\*NLP: Natural Language Processing: refers to the ability of machines to read, understand, categorize, summarize, extract information from and create information in written natural language.



# Example #1: A.I.-enabled work item processing

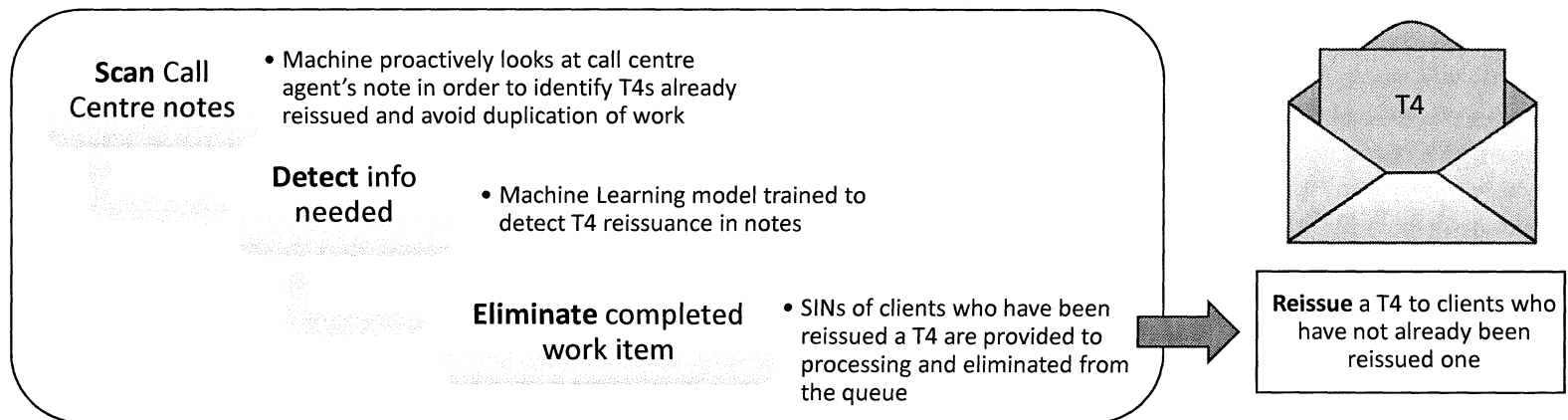
## Business Context

Canada Pension Plan (CPP) and Old Age Security (OAS) recipients receive annually a T4 form containing the benefits received that they will need to report on their tax return. Many T4s are returned to Service Canada due to address changes or other reasons and must be investigated.

## Problem

Each year, large numbers of clients follow up with Service Canada to request a duplicate T4, which is immediately reissued. However, the reissuance is only captured in call centre notes and the resource intensive investigation of every single individual returned T4 is still performed.

## Solution



## Value

- Low-value work eliminated: 50,000 work items, or 2.5 FTEs, saved annually, while development only used 0.1 FTE
- Work inventory is reduced, with faster processing for remaining clients
- Once built, the models can be repurposed to answer new questions



Employment and  
Social Development Canada

Emploi et  
Développement social Canada



# Example #2: A.I.-enabled media monitoring tool

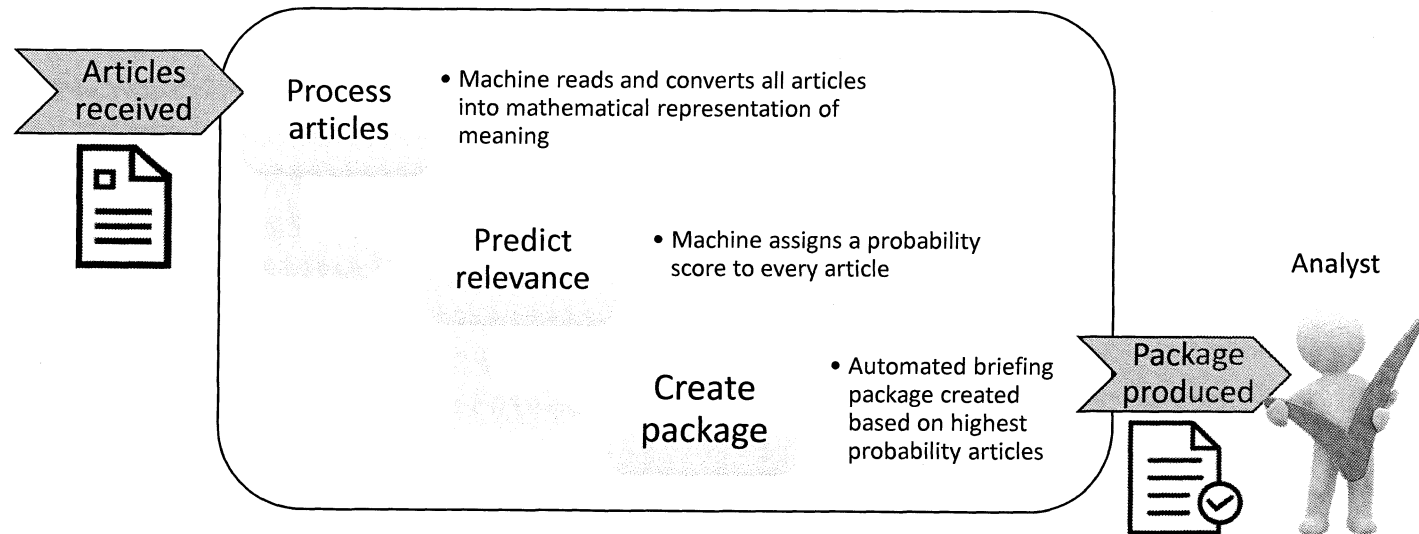
## Business Context

Every day, a package is produced for the ministers that provides pertinent news articles pertaining to ESDC's mandate.

## Problem

The package is produced by analysts who manually review news media starting at 3am each day. The volume to be monitored is often beyond the capacity to review every article and leaves little time for analysis.

## Solution



## Value

- Reduced workload for employees
- Timelier package production allows for real analysis on what the news means for the Department
- More media can be searched and evaluated in a consistent fashion



Employment and  
Social Development Canada

Emploi et  
Développement social Canada

# Example #3: HR Screening Project

## Business Context

Large hiring process such as PM-01 can attract up to 5,000 applicants. Currently, the only way to screen through large volume of applications is to apply filters and to go through the documentation manually. The scope of such screening processes is such that not enough employees can do them all in a reasonable amount of time. This situation is straining resources, its costly and can often lead departments to miss quality candidates.

## Problem

Applying A.I. models to this problem would likely ease the burden on HR and would allow departments to identify and hire quality candidates faster. However, using A.I. to solve this problem poses a number of administrative and legal questions such as *who owns the data ? Can the data be used under the current notice and rules? What is the process for ESDC to gain access to the data?*

## Solution

Setting the projects objectives

- ESDC and PSC working together to establish the objectives and deliverables for this project.

Consultations with key stakeholders

- Consultations with legal services and privacy to make sure ESDC can use data to train models and provide research insights on bias and hiring processes.

Research project

- CDO undertakes the research project to establish if HR screening can be done with advanced analytics tools and to identify bias.

## Value

- Allows the CDO to establish processes and gain knowledge on how to use sensitive data
- Demystify A.I. with legal services and privacy for future projects
- Establish the feasibility of using A.I. to identify and mitigate bias in hiring processes
- Establish the feasibility of using A.I. to improve hiring processes



Employment and  
Social Development Canada

Emploi et  
Développement social Canada

# Example #4: Onboarding Chatbot

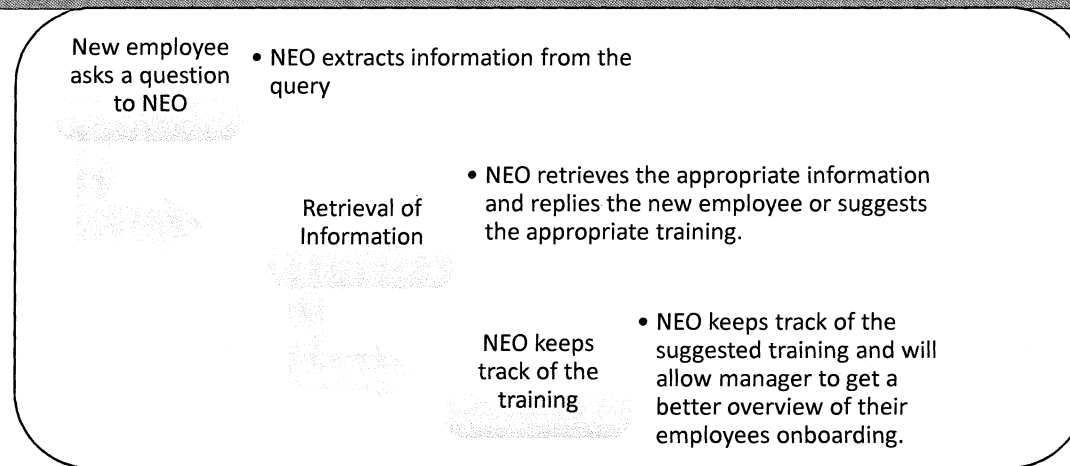
## Business Context

ESDC is large department that onboards many new employees on an ongoing basis, and the process requires many partners (e.g. security, IITBD, HR, manager). Each new employee must obtain several accounts (e.g. for learning, pay, HR), complete mandatory training, learn about the available tools, like National Service Desk, get access to resources needed for their work, and situate themselves in their new work context.

## Problem

Current tools for onboarding include a checklist, email templates, and the creation and implementation of a plan by the manager. This involves manual checks that mandatory training has been completed, requesting certain accounts, and providing information to the employee about resources and guidelines, on top of introduction to work tasks. In many cases the success of onboarding is determined by a manger's ability to identify additional training, information or orientation needed, which creates demands on the manager's time with conflicting priorities and resulting in an inconsistent onboarding approach across the department.

## Solution



## Value

- Faster and more efficient way to onboard employees
- Will allow managers to track the onboarding progress of their employees
- Provides useful information to ESDC on chatbots



# Lessons Learned from the pilot projects – Partnership Matters

- Solve real business problems; do not innovate for the sake of innovation.
  - Every pilot project solves a real business problem and delivers immediate business value for the organization.
- Data scientists and business experts work as a coordinated team
  - Business needs to invest as much time in the development of A.I solutions as the technical experts.
- Internal technical expertise is essential
  - Internal capacity is needed to understand what is feasible and to negotiate with vendors in order to make the most of taxpayer money, understand external solutions and address legal and ethical considerations.
- The underlying data is key
  - The quality of any model relies on the quality of the data and its ability to accurately depict the business process.
- Buy vs. Build; We will do both
  - The A.I. space is not mature so significant customization is often required. There is no out of the box magic A.I. solution that does not require significant investment from the business to implement.

# Partnering Across GoC

- Internally developed A.I. solutions could be repurposed to deliver value to other departments.
- No processes currently exist to scale internal open source development GoC wide.
- The CDO is exploring various collaboration and funding approaches.
- Three projects have attracted significant interest from other departments:
  - Risk Insight tool (Audit) *(Details on the project on slide 16)*
  - Media monitoring tool (PASRB) *(Details on the project on slide 22)*
  - HR application screening tool (HRSB) *(Details on the project on slide 23)*

# ESDC's A.I. Strategy

- The need for an A.I. Strategy:
  - Clarify thinking around challenges and opportunities
  - Support GoC-wide A.I. policies
  - Educate and dispel myths
  - Help ensure responsible use of A.I.
- Grounded in experience
- Describes domains of A.I. investment and applications in ESDC
- A.I. Policy addresses key aspects, including:
  - Legal, Ethics, Transparency, Accountability, Privacy, and Security





# ESDC's A.I. Strategy Principles

1. Develop a modern AI suite to transform the way ESDC delivers services to Canadians
2. Develop a policy for acceptable AI use in light of the risks it poses
3. Strengthen our internal capacity in AI development
4. Organize ourselves to properly steward the most important component of the current AI wave: the data
5. Obtain maximum public value for our investments in AI in discussions with vendors



## ESDC's A.I. Strategy Principles

6. Engage across the organization to promote AI and coordinate initiatives
7. Put in place the right platform for the development and deployment of AI solutions
8. Develop processes and controls for AI models to ensure they do what we want them to do
9. Design a framework for monitoring performance and evaluating success of AI solutions to prove value to Canadians



# Going forward

- Finalize the **Artificial Intelligence Strategy**
  - Version 1.0 of the A.I. Strategy is currently being circulated for comments
- Work with partners across the organization to **build organizational capacity**
- Operationalize models in **production environments**
- **Partner** with other GoC organizations
  - Work with external partners to develop and deploy models that can impact many GoC organizations, train and educate employees and provide advices
  - Leverage other department's proposal to fund the CDO's work to augment the CDO's capacity
- Support **Service Transformation Plan** activities
  - Regular participation in the acceleration hub and advising on specific initiatives
- Establish **Analytics Program**



Stakeholders	Projects
<p><b>Income Security and Social Development Branch (ISSD) (CPPD)</b></p> <p><i>(Project Cancelled)</i></p>	<p><b>Business Context:</b> Canada Pension Plan (CPP) provides disability benefits to people who have made enough contributions to the CPP and who are disabled and cannot work at any job on a regular basis. Medical adjudicators (MA) determine, after reviewing your application, your eligibility.</p> <p><b>Current Situation:</b> CPPD applications have to wait for a medical adjudicator's (MA) review to determine if they require additional documentation. Often this review takes place after several months making it hard to meet the service standard of 120 days.</p> <p><b>Solution Provided:</b> Predicting which client files require additional documentation at receipt to reduce delays and support better and more consistent decisions by CPP-D agents to eliminate costly appeals of decisions.</p>
<p><b>Income Security and Social Development Branch (ISSD) (Homelessness Strategy)</b></p> <p><i>(Project Completed)</i></p>	<p><b>Business Context:</b> ESDC runs, year after year, numerous consultations which almost always contain at least one open ended section where participants can express themselves on a specific topic. In 2017, ESDC ran a consultation around the theme Homelessness. As part of this consultation, the department received hundreds of submissions in free text.</p> <p><b>Current Situation:</b> To get insight in the Homelessness Strategy, analysts would read through all submissions and try to extract key information manually. This method is resource intensive, inconsistent and not reproducible.</p> <p><b>Solution Provided:</b> Through topic modelling, the CDO provided with a fast, scalable and efficient way to extract key information.</p>

Stakeholders	Projects
<p><b>Labour Program – Information Retrieval System</b></p> <p><i>(Project Completed)</i></p>	<p><b>Business Context:</b> About 1 million employees in Canada work in federally regulated occupations. The Labour Program promotes cooperation and fairness and provides expert advice and assistance on labour relations matters to workplaces within the federal jurisdiction. To that intent, the Labour Program has developed numerous services, measures and initiatives to assist employers and employees in creating and maintaining a workplace that is conducive to good industrial relations.</p> <p><b>Current Situation:</b> The labour program, upon requests on collective agreements related to federal jurisdiction occupations, need to extract information manually from a sample of collective agreements because of the large number of agreements.</p> <p><b>Solution Provided:</b> Creating a question and answer tool to categorize and facilitate responses to queries of collective agreements. Model allows for a faster, more efficient and reproducible way of querying collective agreements. The model required 0.2 FTE to create and can be reused over and over at almost no costs.</p>
<p><b>Legal Services (Justice Canada)</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b> External A.I. tools are more and more considered by different branches in their attempt to leverage their data assets. Legal is facing an ever growing volume of documentation and need to keep up with the private sector. However, the knowledge of the different techniques in the department is sparse.</p> <p><b>Current Situation:</b> Legal is interested in acquiring an A.I. solution to keep up with the ever increasing amount of documentation. Legal lacks the proper knowledge to define their requirements in terms of A.I. to external vendors.</p> <p><b>Solution Proposed:</b> Providing technical expertise to advise on Artificial Intelligence projects with external contractors to facilitate the acquisition of the right solution.</p>

Stakeholders	Projects
<p><b>Strategic and Service Policy Branch (SSPB) – Poverty Reduction Strategy</b></p> <p><i>(Project Completed)</i></p>	<p><b>Business Context:</b> ESDC runs, year after year, numerous consultations which almost always contain at least one open ended section where participants can express themselves on a specific topic. In 2017, ESDC ran a consultation around the theme Poverty Reduction. As part of this consultation, the department received hundreds of submissions in free text.</p> <p><b>Current Situation:</b> To get insight in the Poverty Reduction Strategy, analysts would read through all submissions and try to extract key information manually. This method is resource intensive, inconsistent and not reproducible.</p> <p><b>Solution Provided:</b> Automated the extraction of key themes from hundreds of online submissions related to the Poverty Reduction Strategy.</p>
<p><b>Service Transformation Plan (STP)</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b> ESDC is in the midst of transforming the way we deliver services to Canadians. Among the transformation envisioned are modern authentication methods and automated dialoguing solutions.</p> <p><b>Solution Envisioned:</b> Exploring speech recognition technology to identify how it can be used for both client identification as well as analytics.</p>
<p><b>Transformation and Integrated Service Management Branch (TISMB) – OAS related mail)</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b> ESDC is in the midst of transforming the way we deliver services to Canadians. Among the transformation envisioned are modern authentication methods and automated dialoguing solutions.</p> <p><b>Current Situation:</b> A significant amount of emails relating to OAS are hitting the system. Many of those are related to work that cannot be actioned or duplicate work that is currently underway or has been addressed. The current email triaging requires many hours of manual effort, which could otherwise be used in more important areas of processing.</p> <p><b>Solution Provided:</b> Through a number of techniques, the CDO streamlined, closed and aggregated email workflow by triaging emails as they arrived.</p>



Stakeholders	Projects
<p><b>Transformation and Integrated Service Management Branch – OAS overpayment</b></p> <p><i>(Project Completed)</i></p>	<p><b>Business Context:</b> Occasionally under/overpayments are issued to Canadians and Service Canada is notified by the clients. The errors are noted in a system. In order to ensure that every Canadians gets what he/she is entitled too, intensive manual efforts are needed.</p> <p><b>Current Situation:</b> In a number of cases, overpayments and underpayments are issued by ESDC to clients. Information Technology Renewal Delivery System (ITRDS) notes contains information on those over/underpayments. The Transformation and Integrated Service Management Branch employees are manually going through files and extract the information from the ITRDS Notes. This work is time consuming, resource intensive and inconsistent.</p> <p><b>Solution Provided:</b> The CDO used predictive models to flag ITRDS notes reporting an over/underpayment amount over a certain threshold. The model demonstrated accurately that it could extract specific information. The model build by the CDO using only 1/3 FTE is efficient, consistent, is replicable and frees up valuable analysts time. Also, the model developed here can be repurposed to do similar task internally at ESDC or in other departments facing similar issues.</p>
<p><b>Transformation and Integrated Service Management Branch – T4 Processing</b></p> <p><i>(Project Completed)</i></p>	<p><b>Business Context:</b> Each year Canada Pension Plan (CPP) and Old Age Security (OAS) recipients receive an information slip containing the information they will need to report on their tax return. If they are recipients of the CPP the slip is a T4A(P) while it's a T4A (OAS) for OAS.</p> <p><b>Current Situation:</b> Service Canada processing network receives numerous returned T4's due to changes in client's address or for other reasons. A significant number of clients follow-up with Service Canada to request a duplicate tax slip. Processing is not made aware which in return create intensive manual investigation.</p> <p><b>Solution Provided:</b> The model identifies when a client's T4 as being reissued, provided the SIN associated with that T4 to processing and eliminates it from the queue. 50,000 work items, or 2.5 FTEs, saved annually, while development only used 0.1 FTE. Work inventory is reduced, with faster processing for remaining clients. Once built, the models can be repurposed to answer new questions.</p>

Stakeholders	Projects
<p><b>Transformation and Integrated Service Management Branch – Involuntary Separation – GIS</b></p> <p><i>(Project Completed)</i></p>	<p><b>Business Context:</b>            The GIS provides a monthly non-taxable benefit to Old Age Security (OAS) pension recipients who have low income and are living in Canada. In January 2017, guidelines for Guaranteed Income Supplement (GIS) were changed. The GIS guidelines, in the case of a partner admitted to long-term care facility would no longer allow those affected to be assessed as 'single'. This situation leaves one person to pay both household costs and long-term care facility expenses with less GIS support. The Minister of Families, Children and Social Development announced that the guidelines were to be changed to what they were prior to January 2017.</p> <p><b>Current Situation:</b>            After a policy change, a number of applicants were denied GIS. The change in policy was reversed and analysts would need to go through all . Information Technology Renewal Delivery System (ITRDS) notes to identify the clients that were denied.</p> <p><b>Solution Provided:</b>            The CDO built a model to identify those automatically, saving numerous FTEs that can be allocated to other tasks.</p>
<p><b>Innovation, Information and Technology Branch (IITB), Shared Service Canada (SSC) (Data Lake)</b></p> <p><i>(Project Underway)</i></p>	<p><b>Current Situation:</b>            ESDC analysts and external researchers requiring protected data must go through a lengthy and complicated process to access data.</p> <p><b>Solution Proposed:</b>            Exploring a Protected B Cloud Pilot on a variety of data sources including image and speech to provide easier access and advanced analytical tools to researchers.</p>

Stakeholders	Projects
<p><b>Public Affairs and Stakeholder Relations Branch (PASRB) – Media Monitoring (Christine)</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b>            Everyday, thousands of articles and news stories on ESDC or its Ministers are produced. Keeping track of what is being said about the department, its programs or its Ministers with manual techniques is becoming more and more difficult as the speed at which news are produced.</p> <p><b>Current Situation:</b>            The Government of Canada (GoC), through Public Safety Canada, purchases a list of articles pertaining to all federal departments. Those lists are then shared with departments for further analysis. ESDC employees currently come in to work at 3 A.M. in order to open and read between 800 and 1000 articles. Of those, about 100 are evaluated and summarized for ESDC's higher management. Manual monitoring of a sample of all news related to the department is resource intensive and time consuming.</p> <p><b>Solution Proposed:</b>            The CDO is working on a tool that will rate not only a sample, but all the articles provided by Public Safety Canada everyday. It will rate articles according to their content, the mention of an ESDC program or it ESDC's Ministers are mentioned. Once the monitoring part of the model is up and running, an automatic summarizer will be implemented.</p>



Stakeholders	Projects
<p><b>Human Resources Service Branch – Application Screening Automation</b></p> <p><i>(Project Underway)</i></p>	<p><b>Business Context:</b> Screening for selection processes is lengthy and onerous. There are thousands of applicants for entry positions such as PM-01 and AS-01 and too few people to process them.</p> <p><b>Current Situation:</b> Entry positions receive upwards to 5,000 applications which are screened manually. The scope of the screening process is such that not enough employees can do them all in a reasonable amount of time so a random sampling approach is often used. This is straining resources , quite costly, and can often miss quality candidates.</p> <p><b>Solution Provided:</b> The CDO is developing a machine learning tool to can classify candidates based on selection criteria. A classifier will be trained on historical data from past selection processes to screen all applicants. A second model using similarity of answers and content from the curriculum vitae can also be used for ranking and support the predictions of the classifier (ensemble approach). The tool would be deployed as an application that processes the PSRS data and presents results of the screening with descriptive information. Significant attention is being given to the legal and ethical aspects of this tool and is being developed with Legal at the table.</p>
<p><b>Human Resources Service Branch</b></p> <p><i>(Project Completed)</i></p>	<p><b>Business Context:</b> In order to improve employee retention and identify shortcomings in employer-employee relations, HRSB runs an Employee Exit Survey to all employees leaving the department. Some of the fields in this survey are free text.</p> <p><b>Current situation:</b> Extraction of key information would need to be done manually, which is inconsistent and resources intensive.</p> <p><b>Solution Provided:</b> Automated the analysis of the employee exit survey using topic modelling. Common themes that may have otherwise fallen through the cracks were identified (i.e. the link between mental health and housing). Using natural language processing is faster, unbiased, scalable and reusable.</p>

Stakeholders	Projects
<p><b>Human Resources Service Branch</b>  <b>– New Employee Onboarding (NEO) Chatbot</b>  <i>(Project Underway)</i></p>	<p><b>Business Context:</b>  ESDC is large department that onboards many new employees on an ongoing basis, and the process requires many partners (e.g. security, IITBD, HR, manager). Each new employee must obtain several accounts (e.g. for learning, pay, HR), complete mandatory training, learn about the available tools, like National Service Desk, get access to resources needed for their work, and situate themselves in their new work context.</p> <p><b>Current Situation:</b>  Current tools for onboarding include a checklist, email templates, and the creation and implementation of a plan by the manager. This involves manual checks that mandatory training has been completed, requesting certain accounts, and providing information to the employee about resources and guidelines, on top of introduction to work tasks. In many cases the success of onboarding is determined by a manager's ability to identify additional training, information or orientation needed, which creates demands on the manager's time with conflicting priorities and resulting in an inconsistent onboarding approach across the department.</p> <p><b>Solution Provided:</b>  The CDO in partnership with HRSB is developing a ChatBot/Virtual Assistant for New Employee Onboarding (NEO). NEO will guide employees on what accounts they need, where to find them, where to ask for help or get extra information, and track completed training. The deployment of the bot will standardize the onboarding experience, allow for continued support for new employees, create awareness of ESDC activities, learning experiences and services, as well as allow for tracking of tasks and potential for analytics on the onboarding process.</p>

**MAINTENANT ET DEMAIN L'EXCELLENCE DANS TOUT CE QUE NOUS ENTREPRENONS**

2018 SC-CAMS 000049

# Science des données

Bureau de la Dirigeante Principale  
des données, DGPSS





# Le monde des données a changé, ceci signifie des changements pour EDSC également

- **Les attentes des clients ont changées**

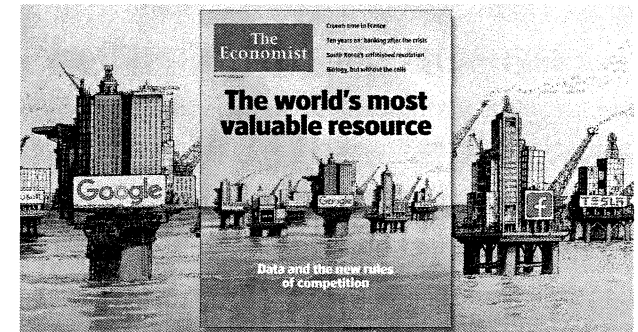
Les clients s'attendent au même niveau de service que celui qui est fourni par le secteur privé, mais qui respecte une norme beaucoup plus élevée de protection de la vie privée. Nous devons intégrer nos données pour fournir un service proactif aux clients dans des environnements sécurisés et gérés.

- **EDSC a changé**

L'accès aux données est nécessaire pour prendre des décisions fondées sur des données probantes, pour produire des résultats et assurer la prestation de services, pour remplir le mandat en matière de transformation des services et pour favoriser la transparence. Nous devons connaître les données à notre disposition et savoir comment les utiliser.

- **La technologie et les techniques d'analyses ont changés**

Nous avons pris du retard dans les investissements sous-jacents nécessaires pour utiliser les données et profiter de leur valeur. Nous avons besoin de personnes, de technologies et d'un programme d'analyse pour rassembler tous les éléments en jeu.



« Je crois que les ministères et organisations du gouvernement doivent se pencher de toute urgence sur cette question. Ils doivent travailler à obtenir les données dont ils ont vraiment besoin pour appuyer leurs activités. Ensuite, ils doivent s'assurer de les gérer correctement et de les tenir à jour. Enfin, ils doivent se servir de ces données non seulement pour informer les activités qui sont au cœur de leurs mandats, mais aussi pour alimenter la reddition de comptes et l'amélioration continue. »

*Rapports du vérificateur général du Canada du printemps 2016, Déclaration d'ouverture, 3 mai 2016*



Science  
des  
données

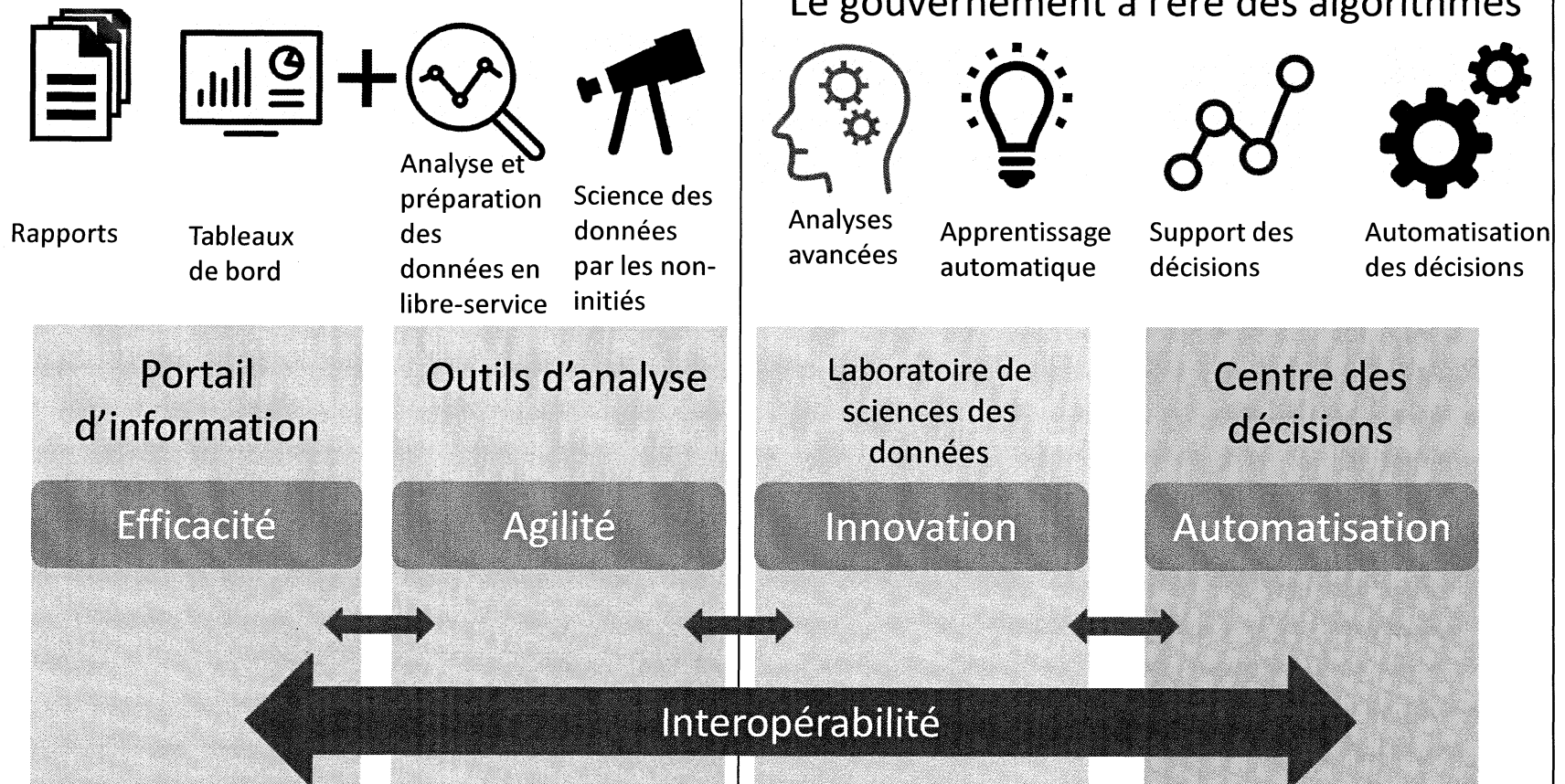
## Pourquoi EDSC doit-il avoir recours à la science des données

- EDSC compte beaucoup de processus manuels
  - Nous avons fait des progrès, mais des changements supplémentaires sont nécessaires pour que les méthodes actuelles de traitement des tâches pour les clients ou pour générer des rapports soient plus rapides, efficaces et qu'elles contiennent moins d'erreurs.
- EDSC dispose d'une grande quantité de données non structurées.
  - La majorité de l'information détenue par EDSC n'est pas utilisée parce qu'il est difficile d'y accéder et de la traiter à l'aide des outils d'analyse.
- La prestation de services d'EDSC est trop souvent réactive et non proactive.
  - Le Ministère doit constamment rattraper le retard sur les arriérés de la charge de travail, ce qui limite les ressources que nous pouvons consacrer à améliorer nos services.

Pour répondre à nos besoins en matière de science des données, le Bureau de la Dirigeante Principale des données d'EDSC élabore un programme visant à développer des capacités analytiques telles que l'apprentissage automatique et l'intelligence artificielle pour découvrir de nouvelles perspectives à partir des données d'EDSC.



# L'éventail de l'analyse



## Questions

- Quels sont les processus opérationnels et TI/facteurs à considérer pour déployer les projets pilotes dans l'ensemble du ministère
- Quelle est la meilleure façon de gérer le changement de culture pour les employés (qui pensent que les machines vont prendre leur emploi) et les clients (qui imaginent une diminution des interactions avec des humains)
- Les membres ont-ils eu des difficultés juridiques, éthiques ou autres à défendre des décisions qui ont été prises à l'aide de l'IA? Comment cela a-t-il été géré et quels ont été les résultats?
- Quelles sont les principaux facteurs à considérer, risques et avantages à acheter/sous-traiter la construction par rapport à la développer la capacité interne (par exemple, la rétention de la propriété intellectuelle, les négociations avec les vendeurs, etc.)?





11

## Exemple du Programme du Travail

Contexte  
opérationnel

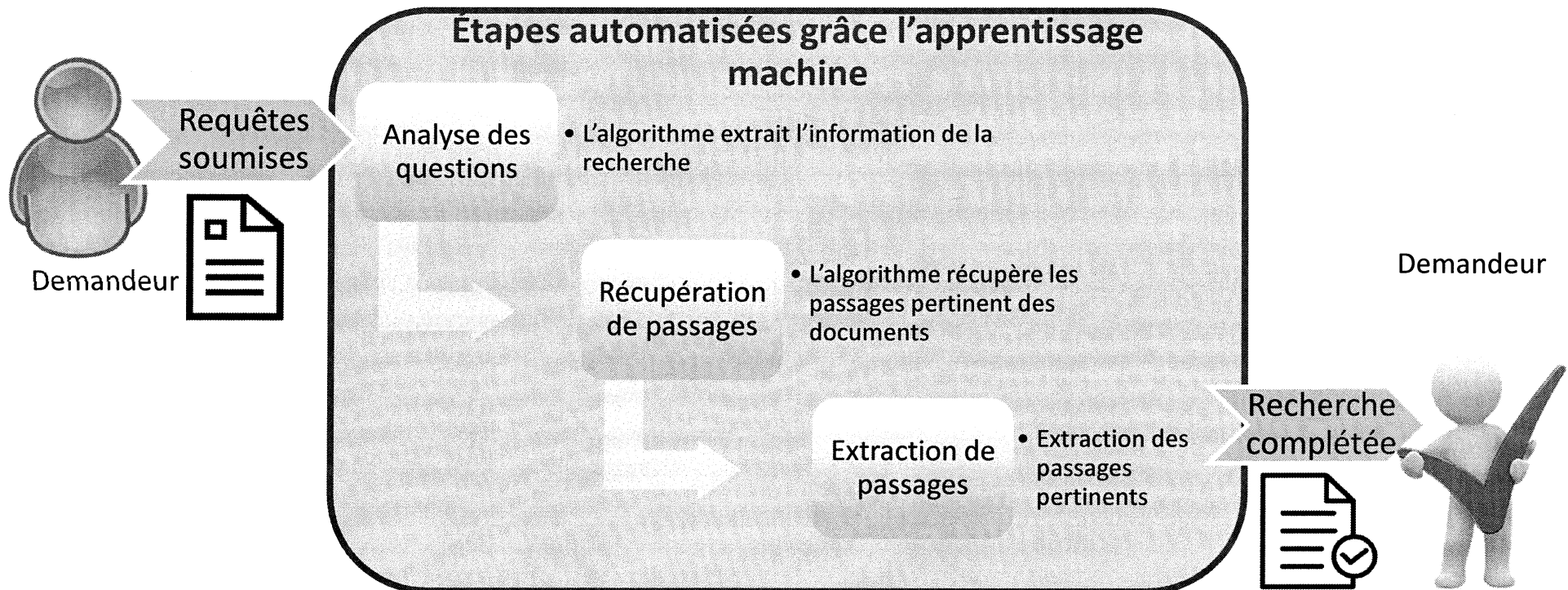
Environ un million de travailleurs au Canada occupent des professions réglementées par le gouvernement fédéral. Le Programme du Travail favorise la coopération et l'équité, fournit des conseils d'expert ainsi que de l'aide sur les questions de relations de travail relevant de juridiction fédérale. À cette fin, le Programme du Travail a élaboré de nombreux services, mesures et initiatives pour aider les employeurs et leurs employés à créer et maintenir un milieu de travail propice à de bonnes relations professionnelles.

Problème

Le Programme du Travail doit souvent extraire manuellement de l'information d'un échantillon des conventions collectives sous juridiction fédérale. Le programme du Travail supervise plus de 50 000 conventions collectives.

Valeur

- Système de recherche plus rapide
- Les recherches sont faites sur l'ensemble des conventions collectives et non pas sur un échantillon
- Le modèle peut être réutilisé sur des problèmes similaires



13

Contexte  
opérationnel

## Automatisation des processus manuels

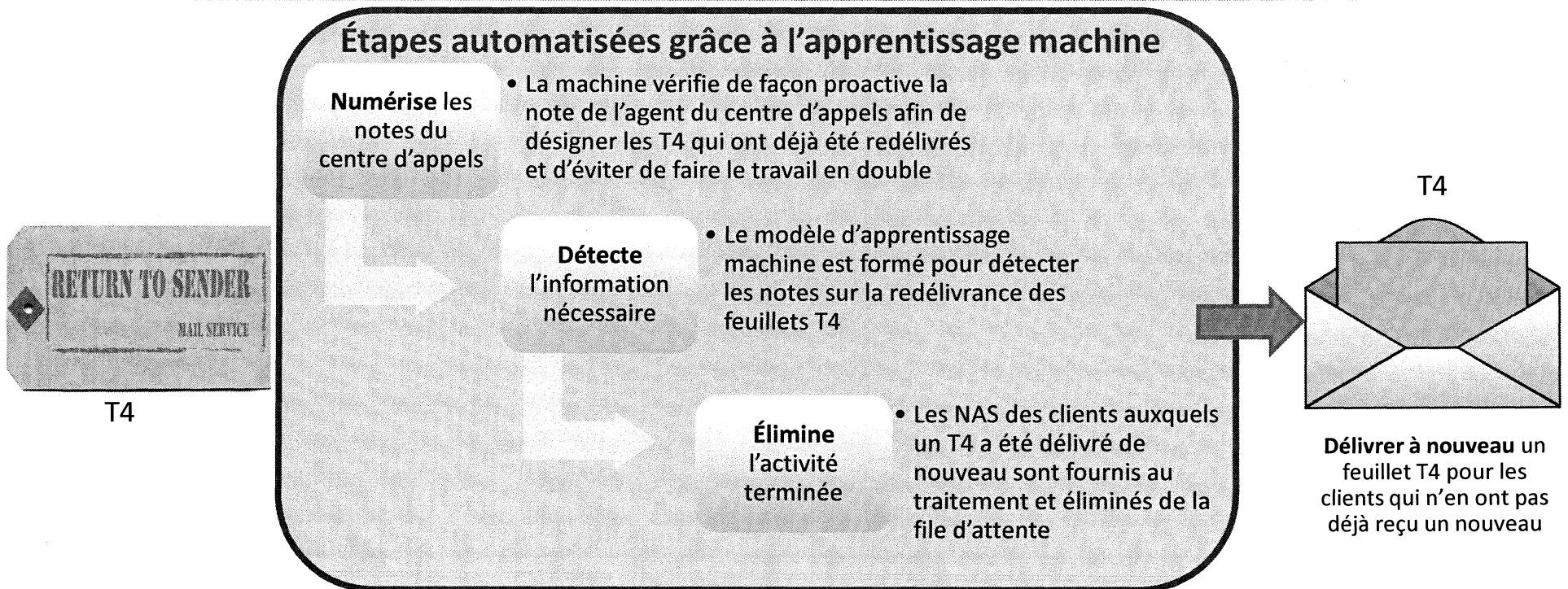
Chaque année, les bénéficiaires du Régime de pensions du Canada (RPC) et de la Sécurité de la vieillesse (SV) reçoivent un feuillet T4 indiquant les prestations reçues qu'ils devront déclarer dans leur déclaration de revenus. S'ils sont bénéficiaires du RPC, le feuillet est un T4A (P), alors que c'est un T4A (SV) pour la SV.

Problème

Chaque année, de nombreux T4 sont retournés au réseau de traitement de Service Canada en raison de changements d'adresse ou pour d'autres raisons. Un grand nombre de clients effectuaient auparavant un suivi auprès de Service Canada pour demander un double du feuillet T4, lequel était immédiatement délivré de nouveau. Le traitement continuait cependant d'effectuer des recherches qui exigeaient beaucoup de ressources pour chaque article retourné.

Valeur

- Travail de faible valeur éliminé : 50 000 articles de travail, soit 2,5 ETP, économisés annuellement, alors que le développement n'utilise que 0,1 ETP.
- Réduction de l'inventaire des activités et accélération du traitement pour les clients restants
- Une fois établis, les modèles peuvent être réadaptés pour répondre à de nouvelles questions





15

## Se préparer aux futures technologies

Contexte  
opérationnel

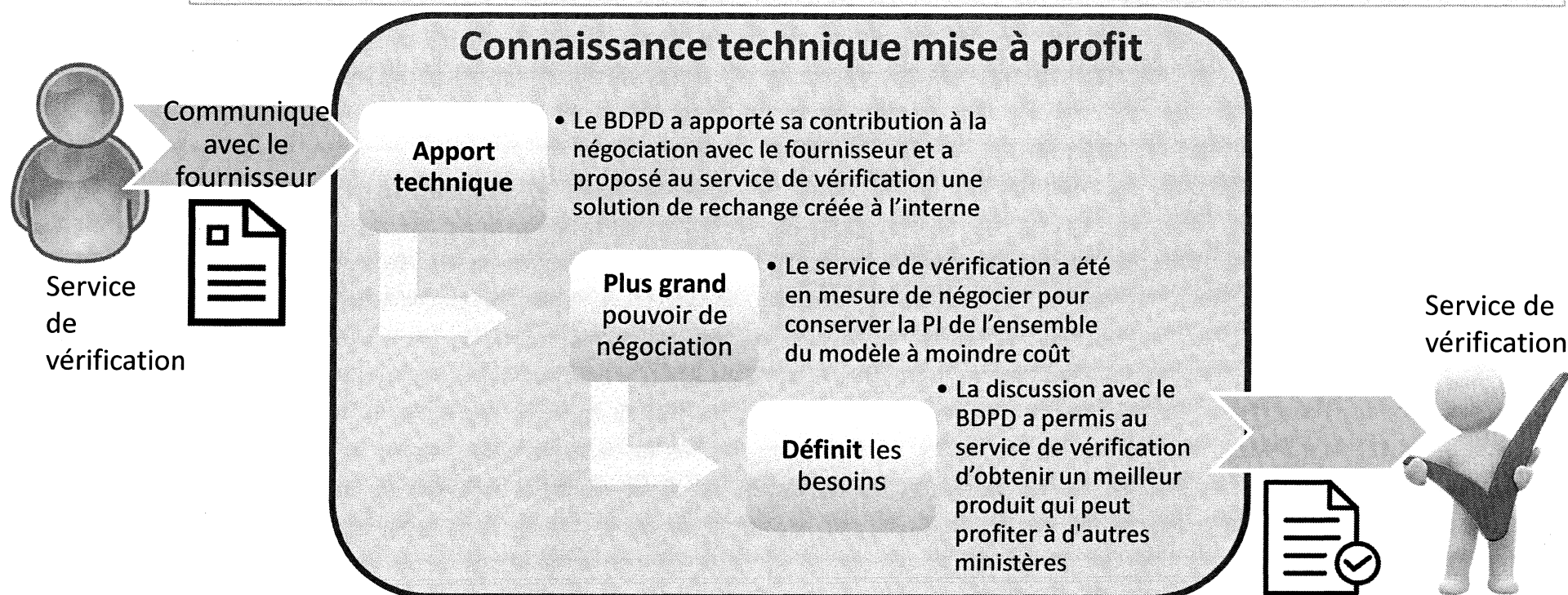
Les outils de l'intelligence artificielle (IA) externes sont de plus en plus souvent considérés par diverses directions générales dans leur tentative d'exploiter leurs données. Toutefois, les connaissances des différentes techniques sont rares au Ministère, de sorte que la possibilité de recourir à des fournisseurs externes est une avenue couramment explorée.

Problème

Le service de vérification a contacté un fournisseur privé pour acquérir un outil de catégorisation de l'apprentissage machine. Le fournisseur privé aurait exigé du Ministère qu'il paie les frais de développement à l'avance, qu'il acquière un abonnement pour utiliser le modèle, et EDSC n'aurait eu aucun droit de propriété intellectuelle (PI).

Valeur

- EDSC garde la propriété intellectuelle, ce qui permet de repenser le modèle, de le réutiliser et de le partager avec d'autres ministères.
- Le service de vérification a obtenu par contrat un meilleur produit au même coût
- L'expertise externe est mise à profit pour respecter des délais serrés



## ANNEXE A : glossaire

- **La science des données** utilise un large éventail de techniques d'analyses sur de grandes quantités de données détaillées pour résoudre des problèmes opérationnels.
- **L'analytique avancée** est très similaire à la science des données, mais se concentre sur les techniques plutôt que sur la fonction globale de résolution de problèmes.
- On entend par « **intelligence artificielle** » la capacité des ordinateurs à accomplir des tâches et à prendre des décisions qui exigent un jugement de niveau humain. L'intelligence artificielle actuelle fait souvent appel à l'apprentissage machine.
- On entend par « **chatbot** » un outil numérique interactif de questions-réponses (robot de discussion).
- On entend par « **apprentissage machine** » les algorithmes informatiques qui sont capables d'apprendre à résoudre des problèmes spécifiques par l'exposition aux données et qui peuvent s'améliorer avec le temps, à mesure que l'on acquiert des données.
- On entend par « **traitement du langage naturel** » les algorithmes informatiques qui s'occupent de l'admission, de l'interprétation, de la synthèse et du discours du langage naturel (écrit et parlé).
- On entend par « **analyse des sentiments** » l'utilisation d'algorithmes pour repérer et extraire la réaction émotionnelle de l'orateur ou de l'auteur à un événement ou à un document.



Partenaires	Projets
<p><b>Direction générale des services des ressources humaines</b></p> <p><i>(Projet terminé)</i></p>	<p><b>Contexte opérationnel :</b>          Afin d'améliorer le maintien en poste des employés et de relever les lacunes dans les relations employeur-employé, la DGSRH mène un sondage auprès de tous les employés qui quittent le Ministère. Certains champs de cette enquête sont en texte libre.</p> <p><b>Situation actuelle :</b>          L'extraction de l'information clé devrait se faire manuellement, ce qui n'est pas uniforme et exige beaucoup de ressources.</p> <p><b>Solution fournie :</b>          Automatisation de l'analyse du sondage sur les départs des employés à l'aide de la modélisation thématique. Les thèmes communs qui auraient pu autrement passer entre les mailles du filet ont été cernés (soit le lien entre la santé mentale et le logement). L'utilisation du traitement du langage naturel est plus rapide, impartiale, évolutive et réutilisable.</p>
<p><b>Direction générale des services de vérification interne (DGSVI)</b></p> <p><i>(Projet terminé)</i></p>	<p><b>Contexte opérationnel :</b>          Les outils d'intelligence artificielle (IA) externes sont de plus en plus souvent considérés par diverses directions générales dans leur tentative d'exploiter leurs données. Toutefois, les connaissances des différentes techniques sont rares au Ministère, de sorte que la possibilité de recourir à des fournisseurs externes est une avenue couramment explorée.</p> <p><b>Situation actuelle :</b>          Le service de vérification a contacté un fournisseur privé pour acquérir un outil de catégorisation par apprentissage machine. Le fournisseur privé aurait exigé du Ministère qu'il paie d'emblée les frais de développement, qu'il acquiert un abonnement pour utiliser le modèle, et EDSC n'aurait eu aucun droit de propriété intellectuelle (PI).</p> <p><b>Solution fournie :</b>          Le BDPD a apporté sa contribution à la négociation avec le fournisseur et a proposé au service de vérification une solution de rechange créée à l'interne. Le service de vérification a pu négocier la négociation de la propriété intellectuelle entière du modèle sans frais d'abonnement permanents. La discussion entre le BDPD et le service de vérification a permis à ce dernier de mieux définir ses besoins.</p>



## Annexe C: Descriptions des projets par partenaires 3/8

Partenaires	Projets
<b>Programme du travail – Système de récupération de l'information</b>  <i>(Projet en cours)</i>	<p><b>Contexte opérationnel :</b>            Environ 1 million d'employés au Canada travaillent dans des professions sous réglementation fédérale. Le Programme du travail favorise la coopération et l'équité et fournit des conseils d'experts et de l'aide en matière de relations de travail aux milieux de travail de compétence fédérale. À cette fin, le Programme du travail a élaboré de nombreux services, mesures et initiatives pour aider les employeurs et les employés à créer et à maintenir un milieu de travail propice à de bonnes relations de travail.</p> <p><b>Situation actuelle :</b>            Le Programme du travail doit extraire manuellement de l'information à partir d'un échantillon de conventions collectives lorsqu'il reçoit des demandes relatives à des conventions collectives liées à des professions relevant de la compétence fédérale.</p> <p><b>Solution fournie :</b>            Créer un outil de questions-réponses pour catégoriser et faciliter les réponses aux demandes de renseignements sur les conventions collectives. Le modèle procure une façon plus rapide, plus efficace et reproductible d'interroger les bases de données sur les conventions collectives. La création du modèle a nécessité 0,2 ETP, et il peut être réutilisé encore et encore, presque sans frais.</p>
<b>Services juridiques (Justice Canada)</b>  <i>(Projet en cours)</i>	<p><b>Contexte opérationnel :</b>            Les outils d'IA externes sont de plus en plus pris en compte par les différentes directions générales dans leur tentative d'optimiser leurs actifs de données. Les services juridiques sont confrontés à un volume toujours croissant de documentation et doivent suivre le rythme du secteur privé. Cependant, les connaissances des différentes techniques sont rares au Ministère .</p> <p><b>Situation actuelle :</b>            Les services juridiques s'intéressent à l'acquisition d'une solution d'IA pour faire face à la quantité croissante de documentation. Les services juridiques n'ont pas les connaissances nécessaires pour définir leurs besoins en matière d'IA et les transmettre aux fournisseurs externes.</p> <p><b>Solution proposée :</b>            Fournir une expertise technique pour conseiller les entrepreneurs externes sur les projets d'intelligence artificielle afin de faciliter l'acquisition de la bonne solution.</p>

## 23 Annexe C: Descriptions des projets par partenaires 5/8

Partenaires	Projets
<p><b>Direction générale de la transformation et de la gestion intégrée des services (DGTGIS) – trop-payés relatifs à la SV</b></p> <p><i>(Projet terminé)</i></p>	<p><b>Contexte opérationnel :</b> À l'occasion, des paiements insuffisants ou en trop sont versés aux Canadiens et Service Canada est avisé par les clients. Les erreurs sont notées dans un système. Pour s'assurer que tous les Canadiens obtiennent ce à quoi ils ont droit, il faut déployer des efforts manuels intensifs.</p> <p><b>Situation actuelle :</b> Dans un certain nombre de cas, des paiements en trop et des paiements insuffisants sont versés par EDSC aux clients. Les notes dans le Système d'exécution du renouvellement de la technologie de l'information (SERTI) contiennent de l'information sur les paiements en trop ou en moins. Les employés de la Direction générale de la transformation et de la gestion intégrée des services examinent manuellement les dossiers et extraient l'information des notes du SERTI. Ce travail prend du temps, exige beaucoup de ressources et manque d'uniformité.</p> <p><b>Solution fournie :</b> Le BDPD a utilisé des modèles prédictifs pour repérer les notes du SERTI signalant un montant de paiement supérieur ou inférieur à un certain seuil. Le modèle a démontré avec précision qu'il pouvait extraire des informations particulières. Le modèle produit par le BDPD avec seulement 1/3 d'ETP est efficace, cohérent, reproductible et libère du temps précieux pour les analystes. De plus, le modèle élaboré ici peut être remanié pour effectuer des tâches semblables à l'interne (à EDSC) ou dans d'autres ministères qui font face à des problèmes semblables.</p>
<p><b>Direction générale de la transformation et de la gestion intégrée des services (DGTGIS) – Traitement des feuillets T4</b></p> <p><i>(Projet terminé)</i></p>	<p><b>Contexte opérationnel :</b> Chaque année, les bénéficiaires du Régime de pensions du Canada (RPC) et de la Sécurité de la vieillesse (SV) reçoivent un feuillet de renseignements contenant l'information qu'ils devront déclarer dans leur déclaration de revenus. S'ils sont bénéficiaires du RPC, le feuillet est un T4A (P), alors que c'est un T4A (SV) pour la SV.</p> <p><b>Situation actuelle :</b> Le réseau de traitement de Service Canada reçoit de nombreux feuillets T4 retournés en raison de changements dans l'adresse du client ou pour d'autres raisons. Un nombre important de clients font un suivi auprès de Service Canada pour demander un double du feuillet d'impôt. Le service de traitement n'est pas informé, ce qui donne lieu à une recherche manuelle intensive.</p> <p><b>Solution fournie :</b> Le modèle détermine quand le feuillet T4 d'un client est délivré à nouveau, fournit le NAS associé à ce T4 au traitement et l'élimine de la file d'attente. Quelque 50 000 éléments de travail, soit 2,5 ETP, sont économisés annuellement, tandis que l'on n'a utilisé que 0,1 ETP pour le développement. L'inventaire de travail est réduit, ce qui permet un traitement plus rapide pour les autres clients. Une fois qu'ils sont mis au point, les modèles peuvent être adaptés pour répondre à de nouvelles questions.</p>

## 25 Annexe C: Descriptions des projets par partenaires 7/8

Partenaires	Projets
<p><b>Direction générale des affaires publiques et des relations avec les intervenants (DGAPRI) – Surveillance des médias</b></p> <p><i>(Projet en cours)</i></p>	<p><b>Contexte opérationnel :</b> Chaque jour, des milliers d'articles et de reportages sur EDSC ou ses ministres sont produits. Il devient de plus en plus difficile de suivre ce qui se dit au sujet du Ministère, de ses programmes ou de ses ministres par le recours à des techniques manuelles, compte tenu de la vitesse à laquelle les nouvelles sont produites.</p> <p><b>Situation actuelle :</b> Le gouvernement du Canada, par l'entremise de Sécurité publique Canada, achète une liste d'articles concernant tous les ministères fédéraux. Ces listes sont ensuite communiquées aux ministères pour analyse plus poussée. Des employés d'EDSC arrivent actuellement au travail à 3 h du matin pour ouvrir et lire entre 800 et 1 000 articles. Une centaine d'entre eux sont évalués et résumés pour la haute direction d'EDSC. La surveillance manuelle d'un échantillon de toutes les nouvelles relatives au Ministère exige beaucoup de ressources et de temps.</p> <p><b>Solution proposée :</b> Le BDPD travaille actuellement à la mise au point d'un outil qui notera non seulement un échantillon, mais tous les articles fournis quotidiennement par Sécurité publique Canada. Il évaluera les articles en fonction de leur contenu, de la mention d'un programme d'EDSC ou de la mention des ministres d'EDSC. Une fois que la partie surveillance du modèle sera opérationnelle, un résumé automatique sera mis en œuvre.</p>


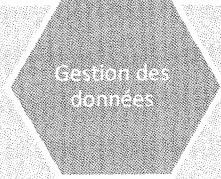



## Annexe D : Objectifs de la stratégie des données

- L'objectif principal de la stratégie de données est de mettre les données entre les mains des personnes qui peuvent générer de la valeur avec le travail qu'elles font.
- Il y a six champs de travail qui permettront de le faire de manière sécurisée, tout en respectant la vie privée des personnes, prouvant que les données peuvent être à la fois plus sécurisées et plus accessibles.
- Deux champs de travail, Accès aux données et Science des données, permettront aux employés et aux partenaires d'EDSC, tels que les membres du Réseau canadien de centres de données de recherche, d'effectuer des analyses et des recherches qui **guideront à la fois nos politiques et nos mandats.**



## ... tout en construisant les fondations de la gestion des données

 <p>Gouvernance des données</p>	<ul style="list-style-type: none"> <li>• Projet pilote de la fonction formalisée de <b>gestion des données</b> avec les RH pour démontrer la valeur d'un processus systématique d'identification et de résolution des problèmes de données.</li> <li>• Projet pilote d'outils améliorés et d'une méthodologie pour évaluer la qualité des données afin d'améliorer la qualité des rapports d'information sur le rendement.</li> <li>• Développer une proposition de métriques de performance pour articuler les résultats attendus de la gouvernance des données.</li> </ul>
 <p>Gestion des données</p>	<ul style="list-style-type: none"> <li>• Proposition d'architecture de référence en matière d'information et d'analyse élaborée avec la DGIIT et approuvée par le Comité d'examen de l'architecture d'entreprise (CEAE).</li> <li>• Plan de transformation des services, point de vue du client, par la Dirigeante principale des données</li> </ul>
 <p>Personnes et habilitation</p>	<ul style="list-style-type: none"> <li>• <b>La Campagne de recrutement postsecondaire</b> pour les volets de données et de recherches est susceptible de déboucher sur un bassin beaucoup plus importants de candidats de grande qualité au niveau maîtrise/doctorat. 120 candidats ont pris part aux entrevues. À titre comparatif, le bassin de 2017 était composé de seulement 15 candidats.</li> <li>• <b>Collaborer avec les intervenants du GC</b> pour faire avancer les initiatives en matière de données (p. Ex., Comité de DG responsables des données, le sous-groupe sur les normes de métadonnées, le forum multipartite sur le gouvernement ouvert, la conférence 2018 sur les données du GC)</li> <li>• Élaborer des plans de communications, d'engagement et de littératie des données.</li> </ul>

Voir l'annexe A pour un aperçu détaillé des étapes importantes



Employment and  
Social Development Canada

Emploi et  
Développement social Canada

Canada

# Employment & Social Development Canada Artificial Intelligence Strategy

## Version 3.0

## Status: Draft

This document presents the Department of Employment and Social Development artificial intelligence strategy developed by the Chief Data office. The strategy details how we will strategically orient ourselves to obtain maximum value for Canadians in our AI investments and sets the stage for AI Policy and Governance to ensure it's used responsibly.

## Table of Contents

Introductory Remarks .....	3
Strategic Initiatives .....	3
1. Why an AI Strategy? .....	4
1.2 What is Artificial Intelligence? .....	4
1.3 AI Strategy Objectives .....	5
2. Grounded in Experience .....	8
2.2 Our Current and Future Priorities for AI Use Cases .....	8
3. Demystifying Artificial Intelligence .....	10
3.1 Communicating ESDC's AI Strategy .....	10
3.2 Background on the Artificial Intelligence Definition .....	10
3.3a Text Classification .....	12
3.3b Information Retrieval .....	14
3.3c Question Answering (for Chat-bots) .....	15
3.3d Text Generation / Document Creation .....	18
3.3e Computer Vision .....	20
3.3f Client Segmentation .....	22
3.3g Strategic Optimization .....	24
4. AI Governance & Policy at ESDC .....	26
4.2 Proposed AI Governance .....	28
Preparing our existing committees for Artificial Intelligence initiatives .....	31
4.3 AI Policy: Key Dimensions of Consideration .....	32
4.3a Legal .....	32
4.3b Privacy .....	34
4.3c Transparency .....	36
4.3d Integrity & Security .....	38
4.3e Bias .....	39
4.3f Impact on the Workforce .....	41
4.3g Performance Measurement .....	42
4.3h Value .....	43
4.3i Other Policy Considerations .....	43
5. The Artificial Intelligence Value Proposition .....	45
5.2 Valuing ESDC's Data .....	45
ESDC's Data Strategy .....	46

5.3 ESDC's Analytics Program: Building for the Future..... 48

    Analytics Program Oversight ..... 49

5.4 The Need for Internal AI Capacity at ESDC..... 51

    ESDC's Machine Learning Seminar ..... 51

5.5 Obtaining Maximum Value with Vendors ..... 52

*Vendor Guiding Principles* ..... 52

6. Links to Other Departmental Initiatives ..... 55

## Introductory Remarks

### Welcome to the Employment and Social Development Artificial Intelligence Strategy Landing Page

Artificial Intelligence is changing the world as we know it, and our clients' expectations and aspirations are changing along with it. This interactive AI Strategy website has been created to highlight how ESDC is approaching AI, both now and into the future.

### Strategic Initiatives

1. Develop a modern AI suite to transform the way ESDC delivers service to Canadians
2. Engage across the organization to promote AI and coordinate initiatives
3. Develop a policy for acceptable AI use in light of the risks it poses
4. Develop effective governance, risk management and control processes for AI models to ensure they do what we want them to
5. Organize ourselves to properly steward the most important component of the current AI wave: the data
6. Strengthen our internal capacity in AI development
7. Ensure maximum public value for our investments when procuring AI technology from vendors
8. Put in place the right platform for development and deployment of AI solutions
9. Design a framework for monitoring performance and evaluating success of AI solutions to prove value to Canadians



## 1. Why an AI Strategy?

Artificial intelligence is receiving an unparalleled level of interest from just about everywhere (including governments, private enterprises, academia and citizens). At ESDC, this interest has manifested itself into several exciting initiatives that are expected to transform how we do business, paired with an equal amount of concern about the risks AI poses. There are questions that need answering related to what activities are underway, what plans are out there, and how we intend to ensure AI is incorporated into the department in a responsible manner that engenders public approval.

The AI Strategy, presented as an evolving, dedicated website, will outlay ESDC's plans, investments and current thinking on artificial intelligence. The Strategy will also provide the foundation to kickstart AI Governance and Policy, which will evolve as the department matures in this area. The Strategy provides an excellent opportunity for departmental staff to become up-to-date in our thinking around AI and enables a myriad of collaborative opportunities to push AI forward as a department to ensure we use it the right way.

### **Ignoring AI is not an option**

One thing that is readily clear is that artificial intelligence is changing the world as you're reading this. From the explosion of chatbots that now represent the first point of contact for many client service journeys, to AI's that diagnose early onset of disease, more and more of our daily life is being augmented by machines. Public expectations are evolving right alongside.

Without substantive investments in artificial intelligence, ESDC's capacity to deliver services will lag further behind that of the private sector (and other governments for that matter), thereby eroding public trust. Service queue sizes, client wait times, client experience/satisfaction, outreach, policy analysis, research, internal services and management practices can all be improved through the use of artificial intelligence, and not embracing these potential benefits would be doing a grand disservice to the taxpayer.

## 1.2 What is Artificial Intelligence?

### **The Strategy Definition of Artificial Intelligence**

For the purposes of this strategy, artificial intelligence will refer to digital solutions that exhibit human or higher-level judgement to carry out tasks for the department.

Further, artificial intelligence solutions must operate in one or more of the following active areas of AI development applicable to ESDC:

- Natural Language Processing
- Computer Vision

- Audio Processing
- Client Segmentation and Advertisement
- Strategic Optimization

Lastly, it is assumed that artificial intelligence solutions covered under the scope of this strategy contain at some machine learning elements that enable them to continually improve their ability to carry out the task.

### Definition Implications

A primary objective of the AI definition is to identify which projects and initiatives will fall under the scope of AI Governance and Policy. The above definition is relatively tight compared to how the term is defined in other areas. Robotic Process Automation, for example, which often does not use machine learning in favour of simpler rule-based methods, is excluded. The objective behind this is to ensure early AI Governance activities focus on core artificial intelligence models that exhibit judgement/discretion in their predictions, as these models represent the greatest unknowns with respect to much of AI Policy. It will be, however, the role of AI Governance to institute a definition of AI that is appropriate with the finalized governance model, and adapt this definition as technology and the department evolve.

Other implications of any AI definition is that it will have inconsistencies with:

- Media (for which the term is often used to incite emotional response from viewership)
- Vendors (for whom there is great incentive to use the term for marketing purposes)
- Our clients (who have a range of views on the subject)

Much of the strategy is dedicated to empowering ESDC with knowledge to have informed conversations about many aspects of artificial intelligence, and many of these concepts apply to solutions that would fall under looser definitions of AI. Sound judgement is required from the whole department with respect to how far prudence outlaid by AI Policy will extend.

Further details about the current nature of AI, and their relevance to ESDC are presented in Section 3 of this strategy.

## 1.3 AI Strategy Objectives

### What is the Strategy Aiming to Do?

The overarching objective of the AI Strategy is to launch official **department-wide** activities that pertain to artificial intelligence. As we begin, these activities will take the form of thorough conversations about the different facets of AI, and the Strategy aims to support this goal by providing sufficient information so that informed discussions can take place. The three main directions this Strategy outlays are the demystification of AI across

the department, the institution of AI Governance and the development of ESDC AI Policy, and positioning ourselves to deliver maximum value for the taxpayer in our AI initiatives.

### **Demystify Artificial Intelligence**

A primary objective of this strategy is to provide an early resource for ESDC staff to gain greater insight into what modern AI is and how it works. Put bluntly, knowledge is power, and this has never been more true than during the information age. As awareness of current AI technology grows, more opportunities for incorporating AI components into ESDC business present themselves.

Further, a comprehensive departmental understanding of artificial intelligence will be critical to build, buy, support, integrate and leverage AI investments in such a way that the utmost care is taken with taxpayer dollars. To support this objective, a department-wide communications strategy for AI activities is being jointly developed by the Chief Data Office and the Innovation Lab.

Section 3 of the strategy presents the strategic direction for AI demystification. Therein, current plans for AI communication are presented. You will also find detailed discussions on different tasks AI can perform, and plain language descriptions of how they work. Also presented are some examples of how different types of machines can be applied in the ESDC context, with the objective of prompting new applications of AI within the department.

### **Set the stage for AI Governance and Policy**

As we integrate AI into our work environments, it is essential that we consider the impacts (both positive and negative) and potential risks before unintended consequences worsen peoples' lives. As the department matures in its use of artificial intelligence, it will accordingly solidify its risk mitigation strategies through appropriate policy.

The development of AI Policy will be achieved by first putting in place a governance framework that includes expertise from the AI-pertinent functions of the department. Further, as many considerations of AI Policy are still in their infancy around the globe, AI Policy will be need to developed through moving AI projects through the governance framework such that specific policy decisions can be informed by real experience. Lastly, AI Governance and Policy will strategically embed itself into existing governing bodies and processes (e.g. Data and Privacy Committee, MPID) where possible, such that new, standalone AI-based committees are kept to the minimum needed.

Section 4 details the strategic plan for an ESDC AI Governance by providing a proposed initial governance model, and some initial timelines for implementation. This section also presents some initial discussions pertaining to different aspects of AI Policy, highlighting the department's current state in these areas, along with anticipated challenges.

### **Position Ourselves to Deliver Value for Canadians**

The ultimate goal of our artificial intelligence activities, as with any public service function, is to deliver maximum value for taxpayer investments. This will be achieved through

making the right internal investments to ensure we're not overly reliant on external providers, organizing ourselves to get better value out of our data, negotiating for vendor services in a way that provides flexibility in how the department can leverage them, and ensuring AI investments result in broadly used solutions that positively impact the lives of Canadians. Section 5 presents strategic initiatives that enshrine taxpayer value as the primary consideration.

### **Intended Audience**

The audience for this document is departmental artificial intelligence stakeholders, including:

- Business and policy areas that seek to make to use of artificial intelligence
- Data science areas that enable AI activities
- Corporate risk management functions that are adapting to the new implications of AI

As mentioned previously, the strategy aims to provide a sufficient starting point such that well-informed conversations can take place. Accordingly, portions of this document will apply to different areas in the department to different degrees, yet the hope is that much of the material is written in an accessible enough manner for broader departmental consumption. Further communication efforts will of course better realize this objective into the future.

### **What the Strategy is not**

The strategy does not go into any detail with respect to financial management or resource allocation, as AI can be incorporated into almost every function in the department.

The Strategy also does not detail specific technologies or tools that form part of the vision, as the rapid evolution of the AI landscape would almost certainly result in a futile exercise.

### **Grounded in Experience**

Finally, the strategy isn't only the result of tireless, deep, policy thinking. The Data Science Division at the Chief Data Office has partnered with a number of different teams within ESDC to deliver on a range of different AI pilot projects. While these projects delivered value for the department in and of themselves, they also were designed to enable us to gain the necessary knowledge to ensure the AI Strategy is informed by proper experience.

Section 2 presents summaries of ESDC AI pilots, and anticipated future use cases that are expected to shape AI Policy as the department matures.

## 2. Grounded in Experience

The machine learning wave of artificial intelligence is relatively new. The level of understanding of recent technological developments vary from organization to organization. The AI vendor landscape is constantly evolving, with the industry not yet mature enough to comprehensively deliver commercial off-the-shelf products. There is no universally agreed upon ethical framework or rulebook for responsible AI development and deployment. In short, the AI issues we're trying to address have not yet been solved globally, and ESDC must put in its fair share of effort to support consensus on these issues.

Prior to and during development of the AI Strategy, it was recognized that our strategic direction needed to be informed by experience through AI pilot projects. These projects have delivered significant value to the organization in their own right, but have also enabled the Chief Data Office and other data science teams in the organization to:

- Dig deeper into the mathematics and algorithms such that inner workings of AI are both accessible and explainable to the department,
- Gain a comprehensive view of relevant existing policy, determine to what degree it applies to AI, and identify where AI-specific policy needs to fill in the gaps,
- Come face-to-face with the risks AI poses, and be able to articulate the challenges ahead,
- Inform the relevant areas of ESDC of what the department will need with respect to oversight, infrastructure and measurement frameworks, and
- Refine our view of what strategies we'll need to put in place and how we'll need to organize ourselves to deliver maximize public value for our AI investments.

The remainder of this section presents some examples of AI use in the department to date, and also presents a strategic roadmap for the types of use cases that the department expects to move into as it matures in its AI understanding.

### 2.2 Our Current and Future Priorities for AI Use Cases

With respect to the future, driving factors of how ESDC will mature with respect to artificial intelligence are the department's current plans related to service transformation and other major initiatives. There are a number of innovative activities underway already that plan to leverage AI, and it's important that efforts related to AI Governance and Policy align such that ESDC is ready to meet the demands when the time comes.

At the same time, as with any new area of investment, a great deal of prudence must be applied to ensure the department does not assume too much risk too soon. ESDC has thus far taken the approach of having its initial investments into AI focus on relatively safe areas that do not have unreasonable levels of downside risk for our clients or employees. However, there is long-term benefit to strategically pushing the boundaries of the

department's comfort level with respect to AI, as greater levels of service delivery can be achieved through this type of innovation.

The following presents a priority list of different applications of AI that represent a roadmap of how ESDC strategically plans to mature in AI, representing increasing levels of risk and reward as we become proficient, and AI Policy takes shape.

**Areas of artificial intelligence that will be addressed by AI Policy in the short-term (2019-2020 - timelines not final):**

- AI's that organize our work and triage workflow
- AI's that help our agents find information related to our programs and policies
- AI's that identify clients in our system with specific characteristics
- AI's that automate low risk business processes
- AI's that scan unstructured client data to populate structured databases
- Internal and external-facing chatbots

We know for certain that these activities will need to be addressed by early AI Policy as they are already being used in our day-to-day work environments.

**Areas of artificial intelligence that will be addressed by AI Policy over the medium-term (Late 2019-2021 - timelines not final):**

Additionally, other AI activities that are being actively investigated to assess potential are:

- AI's that monitor unstructured data for performance reporting
- AI's that inform our targeted outreach operations
- AI's that provide supporting information and recommendations to human agents rendering administrative decisions
- AI's that support policy analysis through emulation of the real world
- AI-based enhanced client authentication

**Items planned for the longer-term:**

Finally, more ambitious initiatives form part of the longer-term vision, but require a more mature AI environment at ESDC than what exists today. Examples include:

- AI's that automatically render administrative decisions in real-time
- AI's that draft documentation and/or communicate directly with clients about their file



### 3. Demystifying Artificial Intelligence

#### 3.1 Communicating ESDC's AI Strategy

Public Servants are the primary audience for the ESDC AI Strategy. They can be broken in to two categories:

1. Public servants in oversight, decision-making, policy and enablement roles that need to understand the policy considerations of developing and deploying AI at ESDC.
2. Public Servants working in business areas that are looking to identify if AI can provide a good tool for dealing with their business problems.

This section of the strategy called 'demystifying AI' provides both audiences an initial understanding of how AI can be used to improve their business processes. The pilot projects highlighted in the draft further illustrate concrete examples of AI in action.

The CDO recognizes that further communications efforts are needed. We will work with the Public Affairs and Stakeholder Relations Branch (PASRB), and the Innovation Lab to create workshops, visuals and other documents so that all audiences can consume and reflect upon the strategy. The more people who are exposed to the strategy the more powerful and meaningful it will be, allowing it to be socialized both inside and outside the organization.

Once we are ready, an external communications plan will also be needed and should illustrate to the public that we are using AI responsibly through sound governance and policy. At this moment, we do not believe active communication to the public is necessary. However, we are always committed to transparency (the strategy has been ATIP'd) and will continue to answer media inquiries on the topic.

#### 3.2 Background on the Artificial Intelligence Definition

Artificial intelligence as a term has no universally agreed upon definition. Its primary use in the most recent wave of excitement is in marketing and advertisement, with many developers having incentive to label their products as AI, even if their level of sophistication would have been state-of-the-art 30 or 40 years ago. Mathematicians, computer scientists, economists, science fiction writers and professionals from a variety of other fields all have different views on what constitutes AI, with some believing that the term should not be used at all due to its ambiguity. One of the benefits of continued use of the term for strategies such as this one, however, is in the level of interest that it generates, which in turn will help move forward the strategic initiatives outlined in this document.

**Artificial Intelligence has a temporal component**

A notion commonly associated with artificial intelligence is that it is present in machines that replicate human judgement. This criterion alone as a definition of AI presents a scope issue, as computers can perform many tasks that are only otherwise doable by humans (for example, adding two large numbers together or identifying whether a word is present in a document). No one would consider these tasks artificial intelligence today, but at one point, they would have been considered the height of computing. It is therefore reasonable to introduce a time dimension into the definition of AI, and we will accordingly restrict our definition to refer to modern applications of machines that replicate human judgement.

Specifically, state-of-the-art research into artificial intelligence is predominantly concentrated in the following areas:

- Natural Language Processing: refers to the ability of machines to read, understand, categorize, summarize, extract information from and create information in written natural language.
- Computer Vision: refers to the ability of machines to classify, recognize patterns in and extract information from images.
- Audio Processing: refers to the ability of machines to listen to, infer sentiment from, extract information from and produce sound.
- Client Segmentation and Advertisement: refers to the ability of machines to analyze and predict patterns of human thinking and behavior, particularly in the area of online decision making.
- Strategic Optimization: refers to the ability of machines to, given a set of possible actions to take, along with a set of constraints defined by the environment in which they operate, make optimal decisions to achieve some objective.

All of these areas are relevant to ESDC, and specific details about each are presented in the remainder of this section.

### **What is machine learning?**

Another term commonly linked with artificial intelligence is machine learning, which has a much more objective definition in the fields of computer science and mathematics. Machine learning refers to a set of statistical algorithms that enable machines to progressively improve performance at a specific task (i.e. to "learn") without being explicitly programmed on how to improve. Machine learning can be further broken down into three main sub-fields:

- Supervised learning, for which the algorithm is presented with available inputs and desired outputs, and programmed to learn itself the best relationship between the two so that it can predict outputs for future inputs.
- Unsupervised learning, for which the algorithm is programmed to find hidden structure in data, without explicitly being told what it is looking for.

- Reinforcement learning, for which the algorithm is not told explicitly what to accomplish, but rather given a reward signal based on its actions, and then has to interact with its environment to determine how to obtain the best rewards.

The current wave of artificial intelligence largely uses machine learning algorithms to develop solutions. Previously, the programming of solutions was much more explicit, leveraging experience of both programmers and relevant field experts. An enlightening example about the current wave of A.I. versus previous waves is perhaps in the analysis of medical scans to diagnose tumors or other diseases/conditions. Twenty years ago, an application developer would have sought the expertise of a top doctor in the field, and explicitly programmed the doctor's thought process in assessing the scan and diagnosing the patient, with some minor probabilistic modeling involved if the doctor assigned a level of uncertainty about any portion of his/her assessment. Today, the machine is given millions of images along with the patients' associated diagnoses by many human doctors, and programmed to determine the relationship (i.e. the "thought process") itself.

### 3.3a Text Classification

#### What is Text Classification?

Text classification is the activity of determining whether free text data (sentences, documents, etc.) meets certain criteria such that labels can be applied to it. Examples of this include whether a given text conveys a positive or negative message, if an email is a request for something or not, or if a given news article is potentially relevant to our department.

Text classification can be very useful to automate the triage and/or labeling of large collections of documents and where there is a constant flow of incoming data. In this context, the automation of the classification can not only free human resources for higher value work, but could also expedite the task such that new opportunities for action are presented.

#### How do Text Classification algorithms work?

Most text classification models today use supervised learning, where a statistical model is able to learn from human-labeled examples of how it should classify documents. For example, a text classifier learns what constitutes a positive message in a document by seeing documents that been previously labeled as positive by a person. The statistical model itself isn't told **how** to determine a positive message from other cases, but will instead use the example pairings of data and human labels to determine what it thinks is the best relationship. This model can then be used to predict/associate categories to future documents that it will have never seen before.

The mechanics behind text classification can be summarized by the following steps:

#### *Estimation and Prediction Process*

- To estimate a classifier, training data needs to be manually labeled with the desired categories.
- Feature development: To enable prediction based on the available text, the machine is given the opportunity to include a wide array of "features" of the text so that it can determine what is important to the classification. Features can include the presence of certain words, the presence of word pairings or triplets, presence of certain character patterns, part-of-speech tags associated with words, and many, many others. This step results in structured data (pertaining to the text) being created from the free text data such that it can be fed into mathematical models.
- In model training, the machine learns the optimal relationship between the available features and the desired categories. In modern text classification, this is achieved using deep learning techniques.
- Once these features are learned and an acceptable error level is reached, the model is ready to infer one or several labels/categories to text examples that it has never seen before.

### **How can Text Classification be used at ESDC?**

Text classification can be applied very widely across ESDC due to the large amount of text data that is received and stored by the department. Examples include:

- Supporting back-office staff by triaging work items that arrive or are internally assigned in written format, enabling more efficient assignment/allocation of work and freeing agents for higher value tasks.
- Undertaking sentiment analysis to categorize the feedback of Service Canada clients as positive or negative.
- Triaging daily news articles, identifying when something is being said about the department, its programs or its Ministers. A text classification tool can be used to rate articles according to their level of importance or relevance to a wide range of professionals across the department.
- Helping with volume management for screening process to assess whether applications meet certain criteria. This could be applied to both internal assessment processes, and public-facing assessment processes.

### **What are some important risk considerations for Text Classification?**

The main risk associated with text classification is that it will classify data incorrectly. Depending on the nature of the classification task, this could result in information not properly being transmitted, work being assigned incorrectly, or improper action being taken on client files. These risks would need to be measured against the error rate of an un-automated process, and the overall value the model provides withstanding these risks.

One useful feature when assessing classification risk is that models give a degree of confidence associated with their predictions, enabling a dimension of risk management (e.g. the automatic triage of the request will only occur if the machine is confident in its prediction above a certain threshold; otherwise the process will revert to the existing manual process). Further, as these models continue to learn from incoming data, AI models will improve their prediction accuracy, providing some assurance that the frequency of incorrect predictions will subside over time.

### 3.3b Information Retrieval

#### What is Information Retrieval?

Information Retrieval (IR) is the activity of obtaining the location of specific information from large knowledge repositories in response to a user need.

In the context of text information retrieval, modern IR mines information from the vast knowledge base and gives results based not just on keywords but also on the intent of the query. Additionally, it can take into account user personal preferences, different meanings of words and spelling errors. Information retrieval is also a critical building block for other AI applications, as it is commonly used in question answering / chatbots to retrieve passages that are likely to contain the answer to a user question.

A few popular applications using IR are the following:

- Search Engines (Google Search, Bing, etc.)
- Job Matching Websites (LinkedIn, Indeed, etc.)
- Google news

#### How do Information Retrieval algorithms work?

The mechanics behind information retrieval can be broken down generally in 3 major steps:

1. Preparing (indexing) the knowledge repository for more efficient information access.
2. Retrieving documents within the repository that match the information need.
3. Ranking the documents retrieved by relevancy to the information need.

For step 1, the purpose of storing an index is to optimize speed and performance in finding relevant documents given a query. Without an index, the search engine would need to scan every document in the corpus, which would require considerable time and computing power. For example, while an index of 10,000 documents can be queried within milliseconds, a sequential scan of every word in 10,000 large documents could take hours.

For step 2, the retrieval of the documents from the corpus is typically done to obtain a smaller set of documents that are relevant to the information need. This is typically useful to filter out irrelevant documents and to reduce the processing required for the ranking step.

For step 3, the purpose of the ranking of the documents is done to be able to assign an order of relevance to each document. The objective is to more highly rank documents that are more likely to contain the information the user is looking for.

### **How can Information Retrieval be used at ESDC?**

IR at ESDC can be used in number of areas. The department holds many knowledge repositories of unstructured data from which it can be difficult to retrieve information (often the solution is navigable drop down menus, which require large amounts of both experience and patience). Also, since IR can be used as a component of other AI techniques, it can enhance other tools by providing relevant information in relation to user needs. The following use cases are possible with IR:

- It can be applied to HR needs in terms of retrieving candidates that match certain requirements.
- It can be applied to recommender systems in retrieving benefits that might interest specific clients based on unstructured data made available.
- It can be used as a component to enhance internal chatbot capabilities.

### **What are some important risk considerations for Information Retrieval?**

Information retrieval generally carries less impact risk than other AI initiatives, as solutions usually provide suggestions to human users with respect to the location of the information they are looking for. Ideally, if the wrong information is ranked higher than the correct information, the user can manage the ranked results to find the particular clause they are looking for. However, ineffective IR tools are not riskless. In more serious cases, the IR tool may not present the correct passage as an option, and if the user is unaware, could proceed with downstream tasks based on incomplete information. IR tools can also be frustrating to use if they are consistently ineffective at providing the user the information they need.

This risk, of course, needs to be measure against potential risks of alternative solutions. Drop-down menu interfaces typically require significant investments in training to become an effective user, and more often than IR tools present the risk that the correct information will not be retrieved, especially for inexperienced users who do not know where to look.

Sources: - [Information Retrieval and Evaluation of the Privacy Risk on Twitter](#)

## **3.3c Question Answering (for Chat-bots)**

### **What is Question Answering?**

Question answering (QA) is a field of research in artificial intelligence that uses machine learning techniques to automatically answer questions posed by users. Questions are typically either typed into a chat window, or spoken aloud and then converted into text via speech recognition software. Responses provided by the machine extend beyond information retrieval (which provides a list of relevant passages from a knowledge



repository that a user can browse), and instead aim to retrieve the direct answer to the question by extracting it from the most pertinent location. The answers are then provided back to the user by presenting response text on the screen, or converted to speech using text-to-speech software.

Examples:

Q: What is the maximum monthly OAS pension one can receive in September 2017? A: \$583.74

Q: Who is eligible to sign an EI Sickness medical certificate? A: A medical doctor or other medical practitioner (health practitioner)

### How do question answering algorithms work?

The mechanics behind classical QA algorithms differ across organizations, but generally follow these major steps: (reference Jurafsky & Martin)

1. Based on the question, determine what type of response the user is looking for (e.g. location, regulation, and date).
2. Generate a query based on the wording of the question and the response type identified in (1).
3. Rank documents and databases in a knowledge repository by relevance to the query generated in (2) using information retrieval techniques, then rank passages in the highest ranked documents again by relevance.
4. Use information extraction techniques to generate a list of possible answers, rank them, and provide the highest ranked response(s) back to the user.

The knowledge repositories used in these algorithms differ based on function. General purpose QA tools use the Internet as the main repository, whereas organization-specific QA tools will use organization-specific documentation sets, which are often proprietary and not available online. Many of the sub-algorithms developed in the major steps listed above are not as general as those for information retrieval, and therefore more work is required to adapt them for other purposes. However, recent progress has been made in this area using modern deep learning techniques, which is enabling models to become more and more effective at successfully answering questions using broad, general-purpose QA databases (reference Squad).

### How can question answering be used at ESDC?

There are several potential applications of question answering tools that could readily provide value to the department, serving both Service Canada clients and ESDC portfolio staff. Examples include:

- A QA tool available to the public that answers general questions related to ESDC program eligibility, application procedures, available benefits and other important

information for which clients would otherwise have to navigate through sub-menus of websites to find the information themselves.

- A QA tool available to support front-line staff during client interactions. The tool would again prevent the need for staff to navigate through list-based knowledge repositories to find the information they need to serve the client, thereby reducing service times.
- A QA tool available to all ESDC staff on the Department's intranet site to support personnel in their day-to-day work. The AI could answer a broad range of questions from "What software is departmentally available for process mapping?" to "How do the CPP child rearing and credit splitting provisions interact?" to "Who is responsible for managing the Job Match algorithm at ESDC?"
- A QA tool to support policy analysts, researchers and program officers for which the knowledge repository is based on reports, data tables and other sources that form the evidence base for their work. Such a tool, for example, would answer questions like "What is the current youth unemployment rate in Chatham?"

The first example, and to some degree the fourth example, utilize public facing information, and could consequently be developed independent of ESDC investment. Google Assistant, as one example, is becoming increasingly effective at general QA over all internet sites, and will naturally increase its capability to answer ESDC related questions as time progresses. The latter two examples include knowledge repositories that are restricted to departmental employees, and would therefore need to be developed internally or through an external contract that provided the repository to the vendor. Additionally, though separate tools are listed above, ESDC could also implement a single multi-purpose QA tool that restricts or expands potential answer types and knowledge repositories based on user type.

### **What are some important risk considerations for question answering?**

The main risk inherent to QA tools is that they will inevitably answer questions incorrectly. This in turn could lead to misinformed clients or staff, who then take inadvisable decisions based on that information. However, competing sources for information retrieval also generate incorrect answers (clients misinterpret information on the web page, or a misinformed staff member passes on incorrect information to another now-misinformed staff member). As time progresses and algorithms continue to improve, it is also inevitable that the likelihood the AI provides the correct information will surpass that of other methods (if not true already).

From a personal information perspective, so long as the knowledge repository does not contain any personal information, there is no privacy risk associated with the associated QA tools. Controls of who can access what components of specific knowledge repositories should be managed independently of question answering front-end applications.

### 3.3d Text Generation / Document Creation

#### What is Text Generation?

Text Generation refers to algorithms in the domain of AI and machine learning that can produce writing in natural language (i.e. they generate their own text). Text Generation is currently in its infancy, but there are already a variety of different real-world applications which provide a glimpse of its potential. Generative algorithms can be used to create texts of arbitrary length, such as a short poem, a description of an image, computer code, or even a full-length novel.

Example: A machine learning algorithm can learn Shakespeare's writing style and begin to generate text that to most observers mimics the language used by Shakespeare. Some compelling examples of the potential of text generation can be found [here](#).

Text generation algorithms are part of a broader family of generative algorithms that are used to generate images, audio, and other media. Many neat examples exist with people [having fun with generative models](#).

#### How does Text Generation work?

There are several different frameworks for generating text (and generative algorithms more broadly). One class involves training algorithms to predict the next word in a sentence, or character in a word. With this type of model, one feeds in input texts, and the algorithm learns to predict the next word or character in a sequence of words or characters based on word/character sequences the machine has observed in large collections of text. Once the model is trained, one can generate text by feeding in input and then having the algorithm predict the next item in the sequence while recursively feeding in the algorithm's output as new input.

Up until recently, a class of deep learning models known as Recurrent Neural Networks (RNNs) were the most effective algorithms at generating text in this fashion. RNNs, and in particular a variant known as Long Short-Term Memory Networks (LSTMs), have the ability to hold previous inputs in memory, making them ideal for a problem in which the task is to predict the next item in a sequence. A drawback of this type of method is that the machine doesn't incorporate any semantic meaning or context associated with the text; it is simply repeating patterns it has previously observed.

An entirely different class of generative algorithms are known as Generative Adversarial Networks (GANs). GANs work by pitting two distinct models against each other, one called the generator, and the other the discriminator, in an adversarial fashion. The basic idea is that the generator network will attempt to generate data that mimics real data (in this case text), while the discriminator network is trained to determine which data is authentic and which data was generated by the generator network. In this way the generator network must learn to fool an increasingly effective discriminator network. For example, one could train a GAN to generate news articles by having the generator network attempt to generate news stories while the discriminator network attempts to distinguish between real news

stories, and the ones that have been generated by the generator network. Both the generator and discriminator network must be machine learning models themselves, so it is natural to use an RNN based architecture for the generator network for the same reasons outlined above.

Most recently, a class of text generation models that have shown an uncanny ability to produce original text use a feature called "attention" (reference paper). Attention models enable machines to maintain context for lengthy passages within text, consequently providing machines the ability to write/speak for longer stretches about a subject. As this is a very active area of research, breakthroughs are happening at a rapid pace, and ESDC must exert the effort required to stay up-to-date, both for the exciting opportunities generative models present, and for the alarming risks they pose.

### **How can text generation algorithms be used at ESDC?**

Text generation has some near-term potential for application at ESDC, as well as significant potential for applications in the medium to longer-term as generative algorithms improve.

In the nearer term, text generation can be used to build more sophisticated chat-bots and to generate summaries of documents. Sophisticated chat-bots, such as the new Google assistant that has emerged from the Google Duplex project, utilize speech-to-text, text-to-speech, and generative models to generate both text, in order to formulate responses, and also audio, in order to generate a voice to converse with the individual who is engaging with the chat-bot. This kind of chat-bot could be used to improve Canadians' experience in obtaining information regarding ESDC's programs and services, as well as providing a way to automate reaching out to Canadians if ESDC needs information, or to notify them that they are eligible to apply for a program and provide information on how to complete the application process.

In the longer-term, assuming that there are significant advances in generative algorithms for text generation, this technique could be used to generate drafts of briefing notes, summaries of documents, and even presentations. There has also been some limited success in having algorithms generate code in multiple programming languages. Generative algorithms have thus far been more successful in the realm of audio and images than text, but it is more difficult to envisage the applicability of these techniques to the department, excepting the aforementioned voice generation for chat-bots.

Another example of how text generation could be useful is to help an algorithm explain the decisions it makes. It is often challenging to interpret and understand why a machine learning algorithm performs the way it does, but adding a generative component to algorithms can allow them to generate text explaining why the model output the way it did, in some ways analogous to human beings explaining their reasoning. This is currently not possible beyond some very limited and rudimentary examples, but there is active research being undertaken.

### **What are some important risk considerations for generative algorithms?**

Generative algorithms pose a serious risk for society as a whole, as they enable individuals to create fake audio, images, video, as well as sophisticated bots online, and, eventually, likely convincing fake news articles. There is also a risk that generative algorithms will start to replicate undesirable behaviour. As with all machine learning algorithms, generative algorithms learn from the data they are trained on, and if poor judgement is reflected in the training data, the algorithm will learn to replicate it. For example, Microsoft released a chat-bot that was trained via interaction with the public, but after conversing with different individuals, the bot started replicating hate speech.

Generative algorithms would also pose some risk internally here at ESDC. These risks include a chat-bot providing incorrect information to Canadians, a generative algorithm producing incoherent content when communicating either internally or with a member of the public, or the reproduction of undesirable behavior as mentioned above. These risks can be mitigated by extensive testing and by limiting the scope in which the algorithms are used to areas in which mistakes will not be overly consequential. Careful controls on the training data used by the generating algorithm are also important to ensure what the algorithm generates remains acceptable.

It is also important to note that if generative algorithms allow us to provide a service that we are not currently able to provide to Canadians, it may not be essential that they achieve a high degree of perfection - as long as they are improving the services we are able to offer, we can tolerate a certain degree of error or imperfection.

## **3.3e Computer Vision**

### **What is Computer Vision?**

Computer vision refers to algorithms that gain a sophisticated understanding of the content of digital images or videos (sequences of images), enabling tasks such as image classification, object detection, scene reconstruction, video tracking, and many others. Input images can take different forms, such as standard 2-dimensional images, image sequences, and 3D medical scan images. The essential element is that digital images have a finite set of digital values (associated with "pixels").

Enormous progress in computer vision has been made worldwide in recent years. Applications include photo-tagging, self-driving cars (for which cameras are constantly recording the vehicle's surroundings to enable it to make decisions), and even art generation. Specific individual applications are very narrow in scope (meaning each model is developed to perform one task in particular), while a general computer vision AI that can completely understand and interpret the full contents of an image to the same degree humans can is beyond current technology. However, at individually targeted tasks, computer vision AI has surpassed human capabilities in a number of areas.

Note that Computer Vision differs from image processing, which is generally associated with image editing and formatting tasks such as image restoration, digital enhancement

and segmentation. Image processing usually results in the manipulation of one image to yield another, while computer vision is more interested in creating structured data from an image (e.g. Is this feature present? Where is it? How big is it?) that can be actioned for other tasks. Both classes of techniques rely heavily on machine learning in modern applications.

### **How do Computer Vision algorithm work?**

Most computer vision applications are developed in a supervised learning context (i.e. they learn from examples). One classical example is a model that is able to recognize if an image contains either a dog or a cat (an example of image classification). To achieve this, the model will be fed millions of pictures of dogs and millions of pictures of cats (labeled accordingly), and teach itself to learn what specific pixel combinations are useful for predicting whether a dog or cat is present. Once trained, the model will be able to make incredibly accurate predictions on new images that it hasn't seen before.

Object detection algorithms, a step beyond image classification, identify where, in an image, a particular object sits. Similar to image classification, most object detection algorithms also use supervised learning, where the model is provided with millions of images along with pixel locations of various objects, and the machine trains itself to localize objects of interest.

Computer vision algorithms typically are provided an image as input, and asked to interpret that image. A traditional pipeline CV applications might look like:

- Image Processing / Manipulation
  - Image Editing
  - Image Enhancement
- Computer Vision Algorithms
  - Image Classification
  - Object Detection

### **How can Computer Vision be used at ESDC?**

As the department still uses physical paper for many processes, there is significant opportunity to use computer vision to render the physical files into machine readable formats and structured database. These applications would primary use an area of computer vision known as Optical Character Recognition (OCR), which aims to retrieve text from images and videos.

Example applications of computer vision in this context are:

- Computer vision can be used to enable the automatic extraction of typed text or hand-written text from documents submitted by clients (e.g. forms, passports, ID cards).
- Computer vision can be used to detect potential fraud attempts in physical documents that the department is receiving when clients apply for benefits.



Examples without text include the use of computer vision to better estimate wait times and queue sizes in Service Canada Centres with video tracking, or for client authentication via facial.

### 3.3f Client Segmentation

#### What is Client Segmentation?

Client segmentation (or market/customer segmentation) refers to dividing clients into groups based on common characteristics so that effective and directed client integration can be taken. For example, resulting segments might identify clients with high needs versus those with low needs, engaged versus disengaged, informed versus uninformed and so forth. In the artificial intelligence context, this activity relies heavily on different types of data/information (e.g. demographics, behaviours, file activity).

Segments can be used for targeted interventions, such as special outreach, directed marketing or policy making, with the aim of intervening before problems arise or identifying groups that are not adequately reached by communications or services. Modern segmentation takes advantage of groups directly identified from the data, which may show unexpected relations between certain groups of people. This is in contrast to 'bought' groups that predetermine which types of people are expected to behave similarly or have similar needs (e.g. youth under 25, urban women with children age 29-45), and may miss more nuanced connections.

In modern AI applications, client segmentation is taken to personalized levels (also known as "segment-of-one"). Using vast amounts of data and deep learning architectures to enable complex structures associated with client behaviours, their likes and their dislikes, AI's are becoming increasingly effective at gaining client attention and prompting response.

#### How does Client Segmentation work?

Data mining techniques identify groups of clients based on factors, traits and behaviours that are measurable in available data. Computational methods are replacing the older "business rules" approach that relies on perceptions of business experts, which can be biased/narrow in focus and unscaleable. The objective of AI client segmentation in the modern context is to identify patterns of behaviour at such a fine level, that predictions about behaviour become different from person to person.

A central theme to modern client segmentation methods is that the data does the talking, almost always using a form of unsupervised learning. Every piece of data available on clients is fed into one or more unsupervised learning algorithms, and the machine is instructed to learn "hidden" patterns:

- For which characteristics or behaviours are groups of clients comparable?
- For which do they differ?
- What types of behaviours correspond to other types of behaviours?

Related techniques such as social network analysis are additionally brought in to help paint the client picture through their known relations. Natural language processing is leveraged for data on the client that is stored in free text format.

Naturally, as the desire of AI segmentation is predict behaviour at levels as close to individual as possible, data pertaining to a massive number of clients are required for the solution to become effective.

### **How can Client Segmentation be used at ESDC?**

ESDC is responsible for a range of services affecting a very large and diverse set of clients across Canada, and segmentation can aid in improved delivery of those services and understanding of our clients:

*Efficiency:* Proactively find unreached clients for communications/services/programs

*Accountability:* Evaluate whether particular needs are met in specific client bases

*Legitimacy:* Understand client groups and their needs in policy-making and program delivery

For example, our *Poverty Reduction Strategy* considers different measures of poverty, since it is not homogenous and relates to different segmentations of people depending on how it is measured. Advanced analytics and segmentation methods can test assumptions about groups, for instance, to find which similarities matter for poverty reduction or which people are out of communication or service coverage with our current policies. Segmentation can help target needs not adequately met because of lack of understanding of characteristics and behaviours, and hopefully result in better service to Canadians in need.

### **What are some important risk considerations for Client Segmentation?**

Certain challenges arise when segmenting individuals for targeted responses, these include:

- The quality of data will be directly reflected in the quality of a segmentation process, mirroring the distribution of individuals in the data. Questions of bias and under-representation in the data must be addressed at the outset.
- Collecting and accessing demographic/location-based/behavioural data (e.g. administrative data, market/satisfaction surveys) is time-consuming and expensive if not already collected for other purposes. Further, the use of all of this data for this purpose may present privacy and/or consent issues.
- Segmentation is inherently discriminatory, since it divides people for specific messaging/engagement. This activity can risk ESDC's reputation if we are not careful to correct bias (e.g. reaching the under-reached) instead of creating further bias. Further, unless properly monitored and appropriate counter-measures put in place, machines may start to segment on traits or behaviours in such a way that it violates the Charter of Rights and Freedoms, Employment Equity legislation or other human

rights protections (for example, segmentation based on skin colour or medical behaviours).

- Finally, many segmentation algorithms are not very interpretable, which often can render the machine's decision difficult to explain.

### 3.3g Strategic Optimization

#### What is Strategic Optimization?

Strategic optimization is an area of artificial intelligence within which machines seek to optimize their actions to most effectively achieve an objective. The most reported on types of strategic optimization AI's developed recently have been machines that play games and machines that trade stocks. In the former case, the most famous game playing machine is Alpha Go (which plays the strategy game Go), developed by Google's DeepMind. Here, the objective is to win the game subject to the games rules, and the machine needs to determine its optimal move set to achieve that objective the most often possible. In the case of trading stocks, the objective would (obviously) be for the machine to make as much money as possible given a specific risk tolerance.

Unlike most applications of natural language processing, computer vision and audio processing, strategic optimization AI's are not necessarily restricted to unstructured data (though they do often use it). Another notable feature that separates strategic optimization from other areas of artificial intelligence is that, in many cases, machines aren't simply capable of replicating human judgement (or the best of human judgement). They are capable of surpassing it, in some cases to a degree which if a human trained his/her entire life, they still could not compete. Alpha Go has no issues dominating top professional human Go players, and high frequency stock trading algorithms have replaced day-traders across the globe.

#### How do strategic optimization algorithms work?

The most successful strategic optimization solutions are primarily developed using reinforcement learning algorithms, though are supplemented by supervised learning to help the machine learn more generally.

Reinforcement learning algorithms can be thought of as repetitive trial and error. The machine will start out by taking effectively random actions (given the situation it finds itself in), and through the reward signal it receives associated with those actions, begin to value taking certain actions over others. After repeated and prolonged exploration of which actions yield the best rewards, the machine will eventually discover which actions are optimal such that it can maximize its objective. Additionally, despite learning continuously that certain actions are superior to others, it will still explore what it believes are sub-optimal actions to reinforce whether or not the action is indeed sub-optimal.

In all real-world reinforcement learning problems, the ability of the machine to explore every possible situation it could find itself in (and learn the optimal action associated with

that situation) is impossible, as there can be an infinite number of such situations. In modern applications, supervised learning algorithms are used to help the machine generalize situations to those it has seen, in order to make hopefully optimal decisions for new situations it encounters.

Often, business problems in which strategic optimization is being applied does not lend itself well to real-world trial and error (e.g. situations where someone's health or financial livelihood may be at stake). In these situations, a simulated environment that is representative of the real-world is created such that the machine can experiment and learn. Once it becomes sufficiently proficient, it is then empowered to take strategic decisions in the real world.

### **How can Strategic Optimization be used at ESDC?**

A field of study that lends itself well to strategic optimization AI is that of Operations Research, where the goal is generally to minimize time (or another resource) dedicated to a set of tasks. Service Canada is one of the largest operational environments in the country, and decisions are made every minute of every day at in-person centres, call-centres and back-office processing centres on how best to allocate resources to minimize client service times. There are numerous opportunities to explore the use of strategic optimization AI's to improve the efficiency of our operations.

Other longer-term potential applications of strategic optimization AI's could include:

- Virtual assistant bots that learn to automate our menial tasks so that we can focus on higher value work.
- The allocation and assignment of computing resources that are shared among users across the department, to minimize both machine downtime and user wait times.
- Malware protection and other IT security functions.
- Program outreach, where the machine learns what is and is not effective in promoting our programs to potential clients.

### **What are some important risk considerations for Strategic Optimization?**

If the machine is given full power to "explore" its options, there is a potential danger in strict trial and error deployments that they model may explore a catastrophic action (e.g. stop payments to all Canadians). In many cases, it is important the machine be given a safe environment to explore its action space, and that the machine's action possibilities and constraints be carefully designed to ensure it operates within desired boundaries.

Relatedly, designing simulation environments can be very resource intensive, and thus cost-benefit analyses are prudent before investing heavily in strategic optimization learning.

## 4. AI Governance & Policy at ESDC

So far, the AI Strategy has articulated the great opportunities that AI presents for ESDC to transform the way we deliver services to Canadians. However, it also presents some very significant risks that need to be managed. Some of these risks are legal in nature; it will be important for ESDC to be clear on what it is allowed to do in the area of AI within the current legislation, and how it needs to protect itself legally. Others are more ethical; the department will need to have a sound understanding of the degree to which Canadians are comfortable with the use of AI in government and how bias will be controlled. Still more are logistical: even with sound AI strategic thinking, how do we ensure our AI solutions perform the way we expect them to, and we maintain control. All of these issues are compounded by the reality that public perception, technological advancements and other facets of AI are evolving extremely rapidly, meaning that answers to these risk considerations are often moving targets. These issues have not yet been resolved globally, so answers will not come easy. However, there are many organizations developing similar policy in this area, so we are not alone.

This section of the AI Strategy outlines how we intend to address these matters through the implementation of *AI Governance* and the development of *ESDC's Artificial Intelligence Policy*. ESDC's AI Policy will be developed through AI governance and will guide decisions in a number of areas related to acceptable use cases, transparency, accountability, privacy, security and model quality, among others, with its ultimate objective being to provide the public with confidence that we are going to use AI responsibly. We are not sure where we will land, so we will need to use pilot projects and a governance framework to figure it out.

The development of ESDC's AI policy will take place over the 2019-2020 Fiscal Year, but several initiatives are helping to lay the groundwork:

- The Treasury Board Secretariat [Directive on Automated Decision Making](#) defines requirements that departments must ensure are met in order to use Automated Decision Support Systems that provide external services. Though systems that provide external services only represent a portion of areas of applicability of AI for ESDC, they do represent the most critical area for AI Policy to protect Canadians. In this respect, the Directive represents a minimum level of care that needs to be applied to use AI. However, for ESDC we'll need more concrete policy to support our AI initiatives and AI Governance framework that extends beyond external services to include internal services.
- The ESDC Data Strategy has initiatives underway focusing on data privacy and security, and many aspects of privacy associated with AI are more related to who/what has access to underlying data (as opposed to what is being done with it). The ESDC Data Policy is currently being developed and will continue to inform the AI strategy.
- AI governance is taking shape with increasingly ambitious pilot projects to establish answers to AI Policy questions. As AI projects move through the governance process we will continue to note key AI policy dimensions that require specific focus. It is

important to keep in mind that the overall goal of the deployment of AI solutions is to improve service for Canadians. AI policies that result in a development process being more cumbersome defeat the purpose. AI policy consequently needs to be grounded in pragmatism, and final policy decisions will be properly field tested with appropriate test projects.

AI Policy will be a critical piece of the AI puzzle, and it's important that it be a joint effort by the department to ensure it meets the needs of all stakeholders. Consultations across the department in the 2018-2019 Fiscal Year have provided useful insights to this updated version of the strategy, many of which are discussed in the AI governance and dimensions of policy considerations.

The CDO will continue to engage across the department to implement AI governance and have an initial AI policy in place by the end of Fiscal Year 2019-2020.

### ***List of Relevant Policy Links***

- [Department of Employment and Social Development Act](#)
- [Treasury Board Secretariat Directive on Automated Decision Making](#)
- [ESDC Data Strategy](#)
- [Draft ESDC Data Policy](#)
- [Old Age Security Act](#)
- [Employment Insurance Act](#)
- [Canada Pension Plan](#)
- [Privacy Act](#)
- [Open as a Foundation for Digital Government](#)
- [Gender-Based Analysis Plus](#)
- [Financial Administration Act](#)
- [EU Guidelines for Trustworthy Artificial Intelligence](#)



## 4.2 Proposed AI Governance

Governance of Artificial intelligence at ESDC is needed in order to establish processes of decision-making among stakeholders involved in developing and implementing AI solutions. The objective of AI governance is to create a system of trust so the department can move forward with developing and procuring valuable AI solutions confidently, knowing that considerations related to privacy, security, ethics, law, transparency, bias and performance are made along the way.

This strategy presents a draft preliminary governance design that will be implemented to guide AI projects. Keeping with the spirit of being grounded in experience, policy will be developed by projects going through the governance process, where consensus will be built over time by decision makers interacting collectively to solve problems. This approach is necessary because AI is a platform technology with many possible applications and various risk profiles; it should be governed with an incremental risk-management approach that is case and context sensitive that is refined as we learn.

The design of the proposed AI governance process below aims to:

- Enable research and experimentation at the outset with low governance barriers, promoting an environment where it is ok to 'fail fast'
- Ensure comprehensive risk mitigation strategies are in place when we move from experimentation to development ("trying" to "doing")
- Leverage and embeds itself within existing governance structures where possible
- Include a lens that emphasizes value to clients, the organization and taxpayers (it's not just about risk aversion)

The design of the governance framework is formatted as a checklist of questions all AI projects will need to answer in order to provide assurance that the solution has been created responsibly. There are three main phases: a research and exploration phase; a design and development phase and an implementation phase. The end of each phase includes a milestone point where review and final decision making takes place by relevant stakeholders. This governance framework is designed to be agnostic with respect to internally developed versus procured solutions, providing flexibility to business areas in how they wish to develop their AI's.

### Research and Exploration

The governance process begins with the objective of enabling the department to openly explore the possible. Business areas are encouraged to innovate and investigate how AI can be applied to a particular set of data to determine if it can improve a business process or solve a problem.

A small check with AI Governance is proposed to:

- Ensure the department has client consent to use their data for this purpose

- Make available tools and algorithms to support the exploration, such that existing work can be leveraged
- Confirm the project is aligned with pre-defined high-level AI ethical principles

After that, it will be up to the respective business lines to determine the feasibility of the solution and whether there is sufficient business value to proceed to solution development.

Phase	Objective	Sample Questions	Who sets the tone?
Initial Governance Check	Quick check to get started	Do we have consent to use this training data for this purpose? Is this type of AI model even feasible given current technology? Has it already been done? Project in line with high-level ethical principles & values?	DPC: Privacy Management Chief Data Office AI Ethics
Research & Exploration	Determine if solution can be built	How might the AI fit into the business process? Is the data of sufficient quality to use for this purpose? Does the data provide the predictability we want?	Business Area (with support from Legal, privacy, CDO, IITB as needed)
Business Review	Should we build solution?	Will the AI provide value to the organization?	Business Area

### Design and Development

During the development stage of an AI project, relevant stakeholders (legal, privacy, CDO, IITB, and CFOB) will be involved so that the business area can keep the AI Policy dimensions of considerations top of mind as they build. Considerations will include model performance, controlling for bias, IT security, privacy, program integrity, accessibility, ethics and compliance with legislation, policy and directives such as the TBS Algorithmic Impact Assessment. This approach will foster a culture of enablement rather than prevention by allowing the business area to bring decision-making functions into the process early on rather than waiting until the review stage to find out that they should have considered something critical. In addition, this approach allows decision makers to make informed decisions about project risks and if the AI solution should proceed to implementation. To ensure we are pushing boundaries an acceptable and healthy failure rate should be established.

Phase	Objective	Sample Questions	Who sets the tone?
AI Design & Development	Design & develop AI solution with stakeholders	How do we measure and control for bias? How do we design the solution while safeguarding assets and information? How do we design the solution to ensure it is compliant with relevant legislation and policy? How do we consider and measure the potential impacts on clients and the dept.?	Business Area (with support from Legal, privacy, CDO, IITB)
	Answer question "Should we use AI for this?"	How will decisions made by the solution be explained? How will the solution uphold program integrity? How will the solution be compliant with the AIA and other TBS standards?	Above AI Stakeholders become involved to help ensure project success
AI Governance Review	Should we implement?	What legal risks are we assuming when using AI for this purpose, and how do we mitigate them? Are there any ethical risks to consider and mitigate? Is model quality acceptable for the business process? Are we comfortable with anticipated model bias and explainability? Does the AI's design meet departmental security standards?	DPC

This phase of the review framework will be embedded in existing governance processes (e.g., MPIB, EARB)

## Implementation and Monitoring

If the decision is made to move the AI solution to implementation, relevant stakeholders will be tasked with making it happen. The objective of this phase is not to determine if the solution will be implemented, but rather how will it be delivered and how will we measure its success over time. With all the considerations made in the development process, the goal is to have close to 0% of projects fail in the implementation phase. The business area will work with the stakeholders to complete all the assurance initiatives needed to control risks and monitor success over-time. IT will play the main role in enabling the solutions deployment by setting up infrastructure, accessibility and data management support. In addition, AI specific issues and monitoring support will be provided as needed by other areas to track changes to the model over-time and protect against reverse engineering.

Phase	Objective	Sample Questions	Who sets the tone?
AI Implementation	How do we deliver the solution?	Will the code for the solution be made open and how? How will data from the solution be collected and stewarded? What infrastructure will support the solution? How will the solution be protected from reverse-engineering? How will the model be peer-reviewed and evaluated? How will the solutions performance be monitored? How will the solution meet accessibility standards?	IITB Business Area Other AI enablers
Review & Monitoring		Is there a monitoring and review cycle in place for the solution? Is there a recourse plan if the solution needs to be changed or removed? Is there a process for reflecting policy or legislative changes?	IITB Business Area Other AI enablers

## Preparing our existing committees for Artificial Intelligence initiatives

A separate committee solely dedicated to supporting AI Policy decisions is not a preferred option, given the existing governance structure in the Department. It is felt that existing governance bodies and structure will be appropriate for AI initiatives and have been highlighted in the governance design:

- The Corporate Management Committee (CMC) is expected to take executive decision on all high-level, contentious AI issues.
- The Data and Privacy Committee (DPC) is expected to take executive decision on most policy aspects related to AI. The DPC, for the foreseeable future, will act as the default body for AI and Data Strategy issues that don't naturally fit with other governance committees.
- The Enterprise Architecture Review Board (EARB) will govern the IT aspects of AI, including architecture and infrastructure governance, setting technology standards and providing IT investment direction.
- Major Projects and investment Board (MPIB) will keep with its mandate of supporting rigorous and transparent project planning, project management, and investment decisions by allowing the AI governance stakeholders to provide assurance in the decision making process.

As an interim measure while AI Policy and Governance are being developed, the Data Science Division at the Chief Data Office will continue offer its services to groups within the department seeking compliance support for its AI initiatives.

### 4.3 AI Policy: Key Dimensions of Consideration

This section presents different ethical dimensions that are expected to be addressed by AI Policy. This does not necessarily represent an exhaustive list, but it is grounded in the experience gained from the AI pilots undertaken to date, consultations and feedback from across the department, and information obtained from thought leaders through research, conferences and forums on artificial intelligence.

We as a department will develop processes and standards for AI as its capabilities advance and change. The development of a comprehensive AI Policy relies on standards created by TBS, across the Government of Canada, in review of policies implemented by other countries (such as the steps taken in the EU to protect and educate citizens about AI and the use of data).

There are many facets to the ethical questions surrounding machine learning and AI. However, AI solutions may often not be worse than current operating procedures in terms of bias, privacy, or security, and have the potential to improve on these aspects. Responsible AI use can create value in tasks that cannot be replicated by hand. AI Policy cannot lose sight of this.

This section of the strategy discusses key components of AI policy; considers the risks and opportunities, and identifies next steps in their development.

#### 4.3a Legal

##### ***Current Requirements and Policy***

All departmental activities (AI included) need to comply with the Department of Employment and Social Development Act, the Old Age Security Act, the Employment Insurance Act, the Canada Pension Plan, and other relevant legislation.

TBS Directives such as the Directive on Automated Decision Making also prescribe actions with which the department will need to be compliant. TBS policy represents a minimum level of care that needs to be applied in the use of AI, and provides a reference point for departmental AI initiatives.

##### ***Considerations and Risks***

## ESDC AI Strategy

DRAFT - V3.0

Three main categories of legal consideration have been identified when the department decides to use AI:

1. **Choice of solution:** It is important that the choice to use an AI solution is appropriate for each associated business problem. For example, if the training of the AI system requires large amounts of data, but too little data is available, or not enough Canadian data that may reflect our laws and unique processes, then the AI solution may not be viable. Legal risks can compound if an inappropriate solution is chosen in the first place, and a "wrong" solution used that results in adverse client decisions is the fault of the department. Put bluntly, "AI for the sake of AI" presents significant legal risk.
2. **Build of the solution:** The accuracy and performance of an AI solution must be appropriate for the task it is assigned to do, so that vulnerable people do not get hurt. Model accuracy levels must be acceptably met prior to model productionization, taking into account the full business context and existing processes. Further, if an AI is feeding into a decision, then the Crown is liable for that decision (the software cannot defend itself). It is critical, therefore, that *ESDC maintain an adequate internal expertise that understands and is able to explain the inner workings of its AI models*. This also has important implications for procured vendor solutions, since vendors do not take on the responsibility of behalf of the department. This has the implication of outright excluding AI solutions that are "black box" for which ESDC staff have limited access and understanding for many business problems.
3. **Delivery of the solution:** Even if a solution is built as accurate as possible for an appropriate business problem, it still must be delivered, maintained and have an appropriate plan for foreseeable roadblocks (such as system downtime). The responsibility requires that we have solutions that can provide correct information but also communicate that information properly. If components of a system are out of commission, can we still explain an individual decision? Has sufficient risk mitigation been properly built in to the AI pipeline from early on in the design stage to post-production system measurement?

### Current Status and Plan

ESDC's Legal Services Branch (LSB) has been very active in the area of artificial intelligence over the past several months:

- An RFI was launched in 2018 to procure private sector expert feedback on what AI can do, key considerations when implementing solutions, and other key considerations.

A pilot that warrants specific mention is a collaborative project between the Human Resources Branch, LSB and the CDO that is exploring the use of artificial intelligence in supporting the initial screening of candidates for a competitive employment process. This project, relative to our other early AI projects, has the potential to impact people's lives to a



greater degree, and was hence chosen as an excellent candidate to explore boundaries and build/refine appropriate AI Policy during project development.

LSB is expected to play a critical role in the development and implementation phases of the AI Governance process. A legal lens is essential for robust AI Policy that protects the interests of the department and its clients.

### 4.3b Privacy

The power of current AI technology is in the data, and almost universally, the more data available the better the AI will perform. However, many concerns exist relating to individual privacy in the collection, storage and use of personal data. AI Policy will need to be reflective of both the relevant legal privacy requirements and government policy objectives, as well as public opinion on appropriate data use, as key requirements for public trust.

#### ***Current Requirements and Policy***

From a privacy perspective, AI Policy will be compliant with the Charter of Rights and Freedoms, the Privacy Act, the Department of Employment and Social Development Act (DESDA), Treasury Board policies and directives and ESDC's policies, specifically, the Departmental Policy on Privacy Management (DPPM). These policies, legislation and directives are the most relevant law pertaining to appropriate collection and use of data for the department. The notion of consent (users being appropriately notified and provided the opportunity to agree to explicit uses of their personal data) features prominently as a key cornerstone of privacy and data policy.

The [2017-2018 Annual Report from the Office of the Privacy Commissioner \(OPC\)](#) highlighted that the office seeks to take an active role with respect to AI being compliant with privacy laws. The OPC has expressed that consent may not always be practicable in the context of AI, where such vast amounts of data are being collected and used for different purposes. Other forms of protection are necessary in such cases, and are actively being investigated. Effective AI governance will enable this investigation by allowing the department to make these decisions through a case and context sensitive approach rather than one size fits all cautionary approach.

An additional OPC guidance piece ([Inappropriate Data Practices](#)), prohibits the following data uses:

- Profiling that leads to human rights law violations
- Uses that are likely to cause significant harm to individuals
- Posting personal information with the intent of charging affected individuals for its removal

These use cases represent downright unethical practices that are extremely distant from the standards to which ESDC holds itself, but are worth noting to highlight the OPC's current concerns with respect to data use.

In October 2018, the International Conference of Data Protection and Privacy Commissioners provided the Declaration on Ethics and Data Protection in Artificial Intelligence. This declaration affirms that the respect of the rights to privacy and data protection are increasingly challenged by the development of artificial intelligence and that this development should be complemented by ethical and human rights considerations.

Much of the privacy legislation and policy that a strong AI strategy will need to encompass are still in the process of being created or reformed around the world. Although the experience that ESDC has gained through its pilot work will undoubtedly prove invaluable, it will be imperative that we keep up to date on the ethical, legal and policy conversations related to privacy for sustaining responsible AI development and implementation over time.

### ***Considerations and Risks***

Many issues related to data and privacy are not AI specific. AI Policy will accordingly need to refer heavily to the broader ESDC Data Strategy and Data Policy on many overlapping issues, and not deviate from the tone set by these pieces.

At some point, with near certainty, a particular AI initiative will face a trade-off between model effectiveness (and increased service/value for the taxpayer) and citizen privacy (via data collection and storage). It will be critical that this trade-off be explored in its entirety and public opinion solicited through consultation and other means. Modern AI at ESDC will be constrained if its related policy and practices are not reflective of modern public views on these issues. This is why ESDC will need to a case- and context-sensitive risk-management approach to identify, monitor and mitigate AI privacy risks.

### ***Current Status and Plan***

The AI governance proposal embeds privacy considerations directly into the process of developing AI solutions regardless who or where in the department they are being created. Via the Data and Privacy Committee (DPC), the department has organized itself to have data more prominently featured in privacy discussion, enabling a more direct conduit to address data- driven AI and privacy issues. AI projects will be presented to DPC, where privacy experts including the Chief Privacy Officer and from the Privacy Management Division will be relied upon to consider the privacy risks of a given project.

Additionally, the Chief Data Office will also actively be investigating and testing privacy-preserving data science options as they become relevant (such as federated learning and homomorphic encryption). These initiatives will allow the department to find attractive points in the utility/privacy trade-off space.

### 4.3c Transparency

ESDC has several responsibilities to the public and other stakeholders with respect to transparency:

- We need to adequately explain decisions that affect clients
- Our business processes need to be subject to independent review/audit to ensure the department is acting responsibly
- ESDC, as part of the federal government family, has expressed an increased desire to be open by default for many of its activities, being mindful of integrity concerns.

These responsibilities will continue to be present as we integrate AI into our systems and operations, and though their overall spirit will not change, the specifics with respect to their implementation will as we develop new approaches.

Further, like almost no field of study before it, artificial intelligence is remarkably open by nature, and this openness is driven by the leaders in the field (Google, Microsoft, Apple, FaceBook, Amazon and IBM). ESDC will need to find its place in this community, noting that it's not making investments at the level of the big players, but still can be a valuable contributor.

#### ***Current Requirements and Policy***

The call for AI models to be interpretable is coming from every area of the department, GoC, academia and the public at large. The TBS [Directive on Automated Decision Making](#) requires that a "meaningful explanation" be provided to affected individuals of how and why a decision was made. Currently, this policy remains at a high level, providing limited description of what is meant by a "meaningful" decision.

Treasury Board has also stressed a number of initiatives related to Government, including the Open First white paper and Digital Playbook. These initiatives represent a new way of thinking for the Government of Canada in an effort to be transparent.

ESDC's current policy is that it does not make open its business processes, for reasons mainly associated with integrity risk.

#### ***Considerations and Risks***

Modern AI models are often "non-interpretable", meaning they are not able to explain the factors that led to specific outputs as have been traditional statistical models (this feature is effectively traded away for powerful problem solving abilities and improved model accuracy).

In cases for which the implications of the decision pose limited risk, this lack of explainability is acceptable (e.g. a model that advances a file in the queue for efficiency reasons). In other cases, the department has a direct duty to explain administrative decisions, and this will not change with the introduction of AI. However, excluding algorithms because they are not readily explainable could decrease the usefulness of an AI

tool. Well thought-out policy pertaining to the level of explainability of models designed for different purposes will be needed accordingly.

Luckily, model explainability is a very active area of research. For example, developing research in XAI aims to decipher the reasons for a decision made by AI, or to create new machine learning methods that are transparent. We are monitoring developments in this area and will investigate how they can be incorporated into our systems.

One intriguing aspect of recent AI research, methodology and use-cases is that *low risk applications* have been made remarkably open and accessible to anyone who is interested. This serves several purposes:

- They make for readily accessible and effective training solutions for AI practitioners.
- Scrutiny can be applied by the open source community to a degree that is effectively impossible in closed environments. Vulnerabilities have the potential to be detected and patched long before being exploited.
- Models, training data and code can be leveraged for other organizations that may directly or indirectly provide benefits to both the department and its clients.

There are, however, potential complications with the full openness of AI methods in our context:

- Fraud and program security
- Given that models will fit into the broader business processes, it may be difficult to make the AI components open given that business processes aren't currently published.

### ***Current Status and Plan***

The Chief Data Office has significant experience with the inner workings of AI models, is expected to play a critical role in aspects of AI Policy that pertain to model interpretability. Business areas will play a central role related to the level of explainability required for respective programs. It is desired that AI Policy get very specific with respect to which AI use cases require different acceptable levels of model explainability.

On the open front, integrity concerns are expected to trump the sentiment for openness with respect to high risk business processes (administrative decisions with significant financial implications being the highest risk). The Transformation and Integrated Service Management Branch will play a critical role with respect to AI transparency policy for administrative decisions. Transparency pertaining to processes with less inherent risk will be a broader discussion.

## 4.3d Integrity & Security

As stewards of approximately \$130 billion dollars of annual benefit payments and the personal data of every Canadian with a Social Insurance Number, the department has every obligation to secure its systems and data. This represents an especially challenging task in the world of artificial intelligence, open source algorithms, API calls and cloud computing.

### **Considerations and Risks**

Although risks associated with ESDC's business processes being compromised have been present since their inception, AI presents new dimensions to these risks. Some potential security threats and issues affecting the integrity of AI and machine learning models are the following:

- Reverse-Engineering [1A]: Computer scientists from Cornell Tech, Swiss Institute EPFL, and the University of North Carolina replicated the output of Amazon AI and BigML models by analyzing a few thousand query-response pairs. The dangers of reverse-engineering are that a black-box can become functionally known and leveraged by attackers, or attackers can implement the stolen model in their own intellectual property without the owner's consent.
- Adversarial Injection Attack [2A, 2B, 2C]: This attack learns how to change the input of a model slightly with the goal of disrupting it and triggering misclassification. This attack can be carried out in such a way that a change to the input is not noticeable via manual inspection, but causes the model to predict a complete different outcome. For example, it's possible by changing just a few of the right pixels in an image of a Quebec driver's license, it will cause the model to predict the presence of an Ontario driver's license card. In order to carry out an injection attack, the attacker requires a reverse-engineered (or the original) model. By using the reverse engineered model, the attacker can algorithmically determine the minimal amount of injection required alter the models outputs. The following are some ways in which an attacker can obtain the information needed to produce successful adversarial injection attacks:
  - Model Parameter Leak: If the black-box model is weights or parameters are leaked (e.g., the weights and architecture of speech recognizing deep neural network), predictions can be generated through exploration of their outputs. In the case of an understandable white-box model's parameters leaking (e.g., a decision tree rules), the attacker will have direct access to the models decision making.
  - Training Data Leaks: If training data is leaked, an attacker could leverage from this data sensitive information revealing vulnerabilities in the machine learning model.
  - Exploratory Attacks Through Indirect Model Access: This is when the attacker will try to understand how the model predicts by testing various inputs and measuring outputs. (e.g., trying various inputs in a web portal and observing the results)

- **Adversarial Dataset Attacks:** The attacker will use adversarial injection attacks to send adversarial data to the system, the goal of which is to have the adversarial data included in the new training data. The training dataset becomes an "adversarial dataset", which if used can corrupt a Machine Learning model. This type of attack is mostly targeted towards models that receive regular feedback. For example, a model that improves a websites search results based on user feedback can be spammed with misleading feedback, the faulty feedback could then be added to the models training data and cause a disruption of the search service. Adversarial dataset attacks can affect both false positives (e.g., deny services to others) and false negatives (e.g., to gain illegitimate access to services).

### ***Current Status and Plan***

As part of the AI governance design, the implementation phase includes explicit considerations on how particular AI solutions will be monitored and reviewed over time, including how they will be protected from third party access and reverse engineering. The Departmental Security Officer and IITB Security lead the security function for automated decision systems in the department, and will play a critical role as security threats take on new forms in the AI environment.

### **4.3e Bias**

Bias is a concern in the development of both policy and human-performed tasks, potentially stemming from cognitive bias or social beliefs, and AI systems conceptually have the ability to overcome these obstacles. However, bias in AI tasks (e.g. decision-making, classification) can also occur depending on the data or algorithm used to create a predictive model, and can have significant consequences when a machine is performing tasks much faster and on a larger scale than a human could. AI bias can come in many forms: data set bias, algorithmic bias, systemic prejudicial bias, and procedural bias, to name a few. It will be critical that ESDC AI solutions exhibit the highest standards of procedural fairness, and consequently minimize undesirable bias to the extent possible.

### ***Bias in training data***

Supervised machine learning requires labelled training data, for instance, records of human-made classifications or decisions related to a set of variables for each (e.g. grant/denial based on an application). If the human-created labels are biased, then the machine can also learn to make classifications with the same bias.

- E.g. Predictive Policing, like PredPol and COMPAS, use historical policing data to train a model for prediction of crime hotspots and likely perpetrators. Due to bias in that data, populations who are historically over-policed or targeted along income and racial lines are disproportionately identified as crime risks by the AI.

### ***Incomplete, skewed or non-representative datasets***

In either supervised or unsupervised methods, if the distribution of categories in the dataset does not adequately reflect the actual distribution, or key variable values are



excluded, then the machine is making predictions on incomplete information and not identifying reliable patterns.

- E.g. Datasets used to benchmark the performance of facial recognition have over-represented male and light-skinned faces by a large majority, thus masking issues in the accuracy of identifying female and darker-skinned faces.

### **Emergent/Similarity bias**

Sometimes bias is intentionally encoded into an algorithm, where the output suggested to a user is meant to be similar to previous searches or personalized based on what the system knows about the user.

- E.g. Facebook News Feed has presented content to users based on what they have already seen, and typically excluded links that take an opposing perspective or on unseen topics, i.e. creating a "bubble", or perpetuating confirmation bias.

The quality of AI, and avoiding bias, require designing the right framework from the beginning, not just checking outputs of developed software. AI Policy should provide sound guidance in this area.

### **Considerations and Risk**

Gaps or historical biases in datasets can cause AI systems to unfairly withhold services, opportunities or resources, which is known as allocative harm. They can also reproduce and amplify harmful stereotypes, causing representational harm. Although technological solutions to reduce biases exist, they are often limited in their capacity to address historical and systemic inequalities. Analyzing datasets for potential biases - and addressing such elements - prior to feeding them into algorithms can help limit unintended harm to individuals and organizations. Some researchers and activists also argue that citizens affected by AI-generated decisions should have the right to see the data, know how it was generated, be able to correct it when necessary and be able to contest decisions.

GBA+ is an analytical process used to assess how diverse groups of women, men and non-binary people may experience policies, programs and initiatives. The "plus" in GBA+ acknowledges that GBA goes beyond biological (sex) and socio-cultural (gender) differences and considers other identity factors, like race, ethnicity, religion, age, and mental or physical disability. As is the case for policies, programs and projects within the Department, the integration of GBA+ in the design, implementation, monitoring and evaluation of AI solutions can help identify and reduce inequalities and bias. For instance, applying GBA+ to ESDC's AI solutions could help identify if certain groups of Canadians are over or under-represented in databases, and, if so, any underlying reasons, and what the potential solutions are.

Another tool, Social Systems Analysis can also be useful to help identify the impacts of AI systems on all parties. Similar to GBA+, it considers the historical, social, political and economic context in which a set of data were produced, including the classification and coding of data. It also examines the extent to which differences in communities' access to information, wealth and basic services shape the data that the AI model is trained on.

### ***Current Status and Plan***

The CDO is expected to take the lead in ensuring model biases are properly measured, understood and addressed from a technical and mathematical perspective project-by-project. Business areas, with other AI stakeholders, will decide what types and magnitudes of bias are acceptable for their business processes.

## **4.3f Impact on the Workforce**

The department highly values its world-class staff, and our current and future efforts into the AI world are to augment and relieve employees, not to replace them. The changing nature of work and the advancement of technology are inevitable, so we want to use AI together with staff in inventive ways that enable service delivery options that were not previously possible.

### ***Considerations and Risks***

To ensure AI activities align with the objective that current and future AI solutions are to augment employees work and not replace them, AI Policy will need to address specific aspects related to that objective:

- AI automation can take over specific tasks, while boosting productivity and, potentially, demand for a service.
- Changes to employee tasks may result in changes to employee work descriptions and trigger changes to classifications and organizational design.
- In time, the public will dictate how best to use AI in its public service, and it is our job to provide the public with options/flexibility.
- The growth in AI is opening up new opportunities in emerging technology leading to new job creation and the need for employees to learn new skills.

### ***Current Status and Plan***

Currently, no AI solutions being developed in the department present a risk of affecting the workforce to the degree of needing organizational change. As part of AI governance, workforce implications will need to be considered as projects are developing, bringing in relevant stakeholders. Stakeholders from the Human Resources Services Branch, labour relations, unions and the Public Service Commission should consider the impacts a particular solution may have on the number of employees needed in a particular process and changes to the nature of their work.

The ESDC Data Strategy emphasizes the need for investing in people and provides training and career paths for all analytics for all analytics personas from policy and business analysts to data scientists.

## 4.3g Performance Measurement

As discussed in Section 3, AI needs to be trained with data to do its job. The training data sets provided to it would establish its sense of right and wrong, which would be defined by the business area and respect the governments policies, legislation and ethical standards. From a technical perspective, the department needs to know that the AI solution is following its training and is performing adequately for its function. Furthermore, AI and machine learning models are not static technologies, so their performance will need to be monitored over time.

### ***Considerations and Risks***

At ESDC, human decision-making is recorded and managed through a structure and process of delegation of authority and is established through legislation such as the Financial Administration Act. These human decisions are assessed through quality assurance and auditing mechanisms to determine if programs, services and internal operations are following roles and responsibilities appropriately. Decisions made using information from an AI solution or made by a solution will also need to be recorded and managed, but this structure of audibility needs to be hardwired into the AI solution or documented by users. Regardless if decisions are made by a human or technology, the department evaluates the outcomes of these decisions to determine if objectives are achieved.

Depending on the business process in which the AI solutions is deployed, the department will need to decide:

- How we will know that an AI solution is making a "right" decision and providing the adequate outputs.
- How the business area will know whether an AI solution is applying the same rigour of analysis as a trained, experienced human agent.
- How often the solution should be audited and if audits should they be transparent to the public.
- How performance and outcomes will be monitored over time.

### ***Current Status and Plan***

The government's current focus on Results and Delivery has required the department to reexamine its business models and processes with a view of achieving better results for Canadians. A key part of this shift is the establishment of timely, complete, accurate and relevant performance information to inform decision makers about programs and services. When investing in AI solutions this same performance examination is needed.

When developing AI solutions key performance indicators (KPIs) should be established to provide a benchmark for success. We cannot use AI just because it is cool. It needs to add long-term value for the citizen. As part of the AI governance design, during the implementation phase the business area will establish KPIs, including ongoing monitoring of data bias. As part of ongoing review and monitoring, ESDC's Evaluation function will not look at an AI solution in and itself but rather the outcomes achieved by its use; while

the process of the decision-making will be what ESDC's Internal Audit will provide assurance around.

### 4.3h Value

Delivering value to department should always be top of mind during all of our business transformation and service improvement initiatives. Proposals to implement an AI solution should always aim to help reduce costs, enhance program integrity, achieve better performance and results, and improve service delivery to Canadians.

Section 5 of this strategy provides an in-depth discussion on the AI value proposition illustrating the importance of valuing our data, building up internal capacity through a robust data analytics program, investing in our people and obtaining maximum value from our vendor arrangements. The AI governance design expresses the need to consider what value a project will bring by allowing the business area to research and explore while bringing in relevant stakeholders to assess return on investment.

### 4.3i Other Policy Considerations

#### *Open Source Software*

Open Source Software (OSS) is commonly used when developing AI solutions. OSS provides value to the department because it can generate an increasingly more diverse scope of design perspective by leveraging knowledge throughout the industry. However, even when freely available, OSS is governed by licensing conditions, which imply strict contractual restrictions. Lack of awareness regarding the existence of such restrictions poses legal risks to the department, both when used in-house and when working with vendors who use OSS in their software development projects.

Some OSS licensing conditions include simple obligations, while other restrictions included in OSS licenses are more complicated and may not provide the intellectual property expectations the department needs. Some conditions include:

- An Infection effect: certain uses of specific OSS can cause the entire resulting software development to be governed by the respective OSS license. The software will be subject to compliance with the very same requirements that were applicable to the OSS component used, turning the resulting solution entirely open.
- Disclosure of complete source code: some OSS includes the obligation to disclose the entire non-OSS source code.
- Commercial distribution not permitted: many OSS licensing policies do not allow users to commercially distribute any deliverable, which includes them. A vendor should not be selling or renting deliverables containing OSS that falls under these licenses.
- License prohibitions: some OSS includes clear license prohibitions to modify or to embed OSS for or within a software deliverable.

To mitigate these risks it is first important for the department to be aware that not all OSS is made equal. It is important for technical staff and management to work together on desired outcomes and conduct an assessment with legal services on the impact of each OSS licensing conditions, including what trade-offs may need to be made to achieve objectives.

### *Other Ethical Considerations*

The considerations discussed above are all important and ethics run through them all. However, an AI solution may be legal, it may not break any privacy principles; it may not pose a high level of bias risk, but it still may not be ethical. Although most technology is designed with the best intentions, it can be difficult to anticipate long-term impacts of a product once it is released and reaches scale.

Effective AI governance and policy can help makers of the technology, project managers, business area experts, engineers, and others get out in front of problems before they happen. AI governance aims to facilitate better product development, faster deployment, and innovation that is more impactful. All while striving to minimize technical and reputational risks.

The department will need to consider the ethical implications of releasing and scaling an AI solution. There are multiple points in the governance design when this should be considered. During the initial governance review, a common sense check should take place as to whether we should proceed or not. The research and exploration phase can also determine that the data and technology will not provide an adequate amount predictability given the decisions we hope to achieve, making it unethical to proceed. In addition, during the design phase we may notice that we will not be able to implement adequate controls to stop the AI solution from causing harm when scaled.

As we develop AI, we should review each project for future scenarios by considering:

- How different users might be affected differently?
- What actions we will take to safeguard privacy, truth in decision-making, democracy, mental health, civic discourse, equality of opportunity, economic stability, or public safety?
- What could we be doing now to get ready for this risky future? Are there any new categories of risk we should pay special attention to now?
- What design, team, or business model choices can actively safeguard users, communities, society, and the organization from future risk?

Recently, a high-level expert group on AI in the European Commission presented their [ethics guidelines for trustworthy artificial intelligence](#). According to the guidelines, trustworthy AI should be: (1) lawful - respecting all applicable laws and regulations; (2) ethical - respecting ethical principles and values; and (3) robust - both from a technical perspective while taking into account its social environment. These guidelines are well suited to help frame our thinking and are in line with our current initiatives.

ESDC needs to align itself with the growing number of voices to ensure that AI is developed in a responsible manner so that we do not lose control. It should be understood, however, that we are (for the foreseeable future) working in the realm of Narrow AI, and humanity has not yet advanced to Artificial General Intelligence.

## 5. The Artificial Intelligence Value Proposition

The AI value proposition is the intention that innovations such as machine learning will improve services to Canadians by speeding up responses to inquiries, benefit delivery and enable greater insight by automating time-consuming internal processes. This will allow more time for reflection, planning and cost-benefit analysis resulting in better policy development.

Achieving this value proposition requires the department to not only invest, but also reinvent our approach to people, processes and technology. Our AI pilot projects have provided us a grounded understanding on how to generate benefits and mitigate risks in order to maximize outputs. This section sets out a vision for how this value can be achieved by considering what will need to be done to shift from a culture of hierarchical and mechanized development to more modular technical design processes. This shift will affect how we build AI solutions in house; how we procure solutions from vendors; and what technology, training and skills are needed both now and in the future. The opportunity for us to set the stage for this shift is ripe because developments in AI are still early.

### 5.2 Valuing ESDC's Data

ESDC's greatest asset when it comes to AI is that the department is a prime generator and user of data. Across the department data generation, management, storage and analysis are fundamental tasks, used to provide benefits and services to individuals and business, and to detect non-compliance, evasion and fraud. However, the department vast data assets have been underutilized in the past.

In order to properly evaluate the investments that ESDC makes in artificial intelligence, the department must first properly assess the value it places on its data and how it perceives them. The desire to change how data is being used at ESDC has led the department to create the position of Chief Data Officer and to establish the CDO office. The mandate of the CDO is to maximize the value of ESDC's data assets, to maximize the way data is collected, stored and analyzed.

- **Collection:** Given a particular problem, what data is needed and what is available? Are their limits to what we are legally authorized to do with the data we do have?
- **Data flows and infrastructure:** How do particular types of data flow through the department and is it reliable? Where is structured and unstructured data stored and who has access?



- Exploring and transforming: Given a particular problem or context, does the data need to be cleaned? Do we have an incomplete data set? Are there other data sets in the department that can be linked together for a more complete picture?
- Data analytics: What are our data stories? Can we use our data to define metrics to track change and our understanding of the data given various factors and contexts? Do we know what we want to predict or learn? Can we create training data by generating labels?
- Learn and optimize: Putting in place AI experiments and pilots to be grounded in experience so that we can learn incrementally, to minimize risk and optimize results.

To achieve the CDO mandate, one of the first tasks of the OCDO was to develop the first ESDC data strategy.

Data is a valuable thing and will last longer than the systems themselves. "- Tim Berners-Lee, inventor of the World Wide Web.

In other words, take care of our data, we never know what problems they can help us solve!

## ESDC's Data Strategy

The vision of the ESDC data strategy is to ensure that employees have access to the data when they need it. To achieve this vision, the BDPD has a number of goals to achieve:

- Make data a business asset through effective governance and stewardship
- Make data accessible and secure
- Transform business group and IT group partnerships to improve data use
- Provide people with knowledge and tools to use data
- Make data science, including the ability to develop artificial intelligences, a core competency of the department.

To achieve these goals, the strategy is based on six pillars: data governance, data access, empowerment, people, data management and data science.

*How will these pillars enable us to put in place an artificial intelligence strategy?*

**Data Governance:** Governance is about answering the following questions: what data do we have; what do we need; where can we find these data; what is their degree of reliability; Who makes the decisions; which rules should be applied?

**Data Management:** Data management is about ensuring the infrastructure is in place to securely store data and to provide users with access to embedded data with the tools they need to analyze it.

For purposes of artificial intelligence, data governance and data management are aimed at ensuring that the data is ready, willing and able to be used once the initiatives are ready for launch. It would be unacceptable for the Department to use faulty data and inappropriate processes.

**Access to data:** Access is about making data available to all who need it quickly and securely while protecting confidentiality

**Empowerment:** Empowerment is empowering people to make better use of data (understanding of data, cultural change, support, communities, communication and tools)

The pillars of "data access" and "empowerment" are aimed at creating an ESDC culture that promotes innovation and experimentation. Such a culture is paramount in all organizations with great aspirations in the field of artificial intelligence.

**People:** The People Component is about recruiting and retaining people with the skills we need, creating the right team structures, and working with ESDC partners.

**Data Science:** Data science involves developing a program to build the analytical capacity to use methods such as machine learning, AI and other methods and tools to discover new information from data analysis.

Data science is the culmination of all the other pillars of the data strategy. In fact, if the data is ready and of good quality, accessible, with the right people present and ready to use it, data science can be effective and can bring the most value to the Department.



The artificial intelligence strategy is an important component of the ESDC analytics program and its value is enabled by a sound data strategy. It defines what people can expect from analytics in the department, the importance of having people with the right knowledge, how to approach the different companies selling artificial intelligence services and infrastructure necessary to optimize the use of public funds.

As can be seen, the ESDC data strategy is dependent on the analytics program while the latter depends on the artificial intelligence strategy. However, it is the latter that benefits most from the data strategy because with the latter in place, all the initiatives presented in the strategy on artificial intelligence can take off.

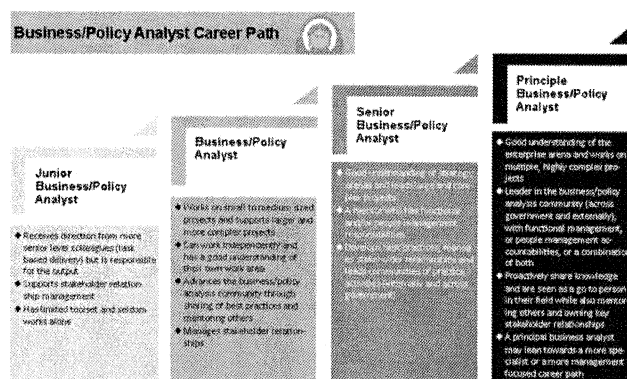
### 5.3 ESDC's Analytics Program: Building for the Future

Robust data analytics is the corner stone of our ability to provide value through AI. This stream of the [Data Strategy](#) is for the department to build upon its people. This includes the need to focus on recruitment and retention of people with the needed skills; putting in place the right team structures; providing the necessary tools; and keeping up to date on industry standards. In the area of Analytics (within which AI resides), this will be realized by the [ESDC Analytics Program](#).

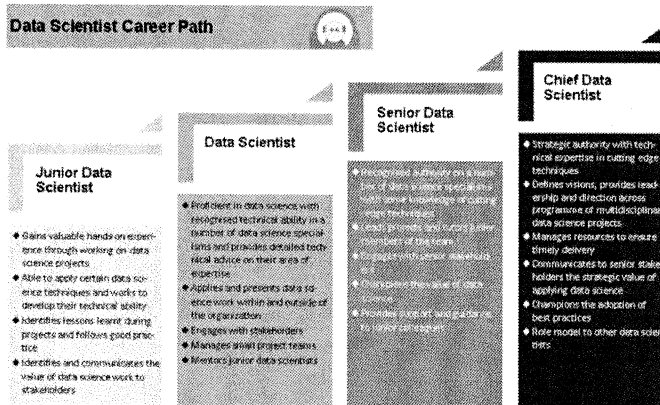
The objective of the Analytics Program is to create a scalable, secure analytics ecosystem that supports all business uses in the department to improve outcomes for citizens, families, organizations and communities. Not all problems that require data analytics need AI to provide useful insights. This is why it is important for the department to have tailored analytics training, awareness and career paths that sustain the right mix of techniques to be applied to different sets of use cases. We must know when AI is the solution to a problem and when it is not. (so we don't ask for a rocket ship, when all we need is a skateboard).

The analytics program establishes training and career paths for all analytics personas. This allows the department to understand and apply the needed roles and responsibilities of all employees involved in the development work and what value each bring to the design process. It enables employees to know what training they need for what role and how it will be beneficial, creating a culture of data literacy continuous learning. In the future this understanding will support the development of workforce adjustments and organizational design shifts brought about by increased automation.

#### Business / Policy Analyst Career Path



## Data Scientist Career Path



## Data Literacy Skills and Learning Objectives

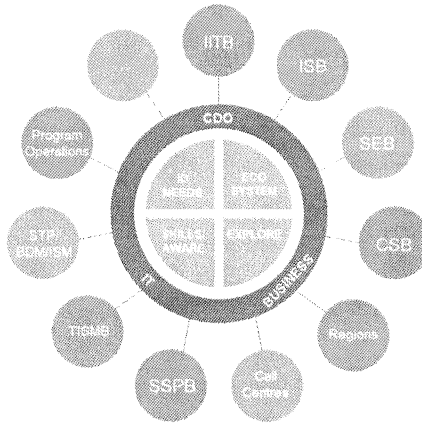
Personas	Data Science			Data Governance		Business Analytics	
	Citizen Data Scientist*	Data Scientist	Data Provisioner**	Data Governor		Analytics Consumer*	Business Policy Analyst
Learning Objectives	<ul style="list-style-type: none"> <li>Develop a depth of knowledge and understanding of principles, tools and applications of data science</li> <li>Apply quantitative modeling and data analysis techniques on economic, social, or service delivery concerns that are relevant to ESDC in order to inform policy and decision-making</li> <li>Gain knowledge of a broad range of quantitative methods such as data mining, predictive modeling, machine learning, and statistical intelligence to conduct data analysis of business problems</li> <li>Apply a variety of data management techniques to optimize business value of data at important points in the data life cycle, from collection and storage to dissemination and integration</li> <li>Effectively communicate findings and present results from data analysis using data visualization techniques</li> </ul>			<ul style="list-style-type: none"> <li>Develop the knowledge and understanding of data governance framework and processes, and the processes and the disciplines for managing data and information</li> <li>Be able to define and understand the roles and responsibilities associated with data stewardship and other roles within the data governance framework, and to ensure clear accountability</li> <li>Develop and implement a framework to ensure the quality of departmental data to effectively improve overall data strategy</li> <li>Obtain the necessary technical knowledge and skills to build, implement, maintain, and improve data security safeguards, build quality and data architecture and data, and implement data protection standards</li> <li>Define, establish, and report on metrics to evaluate the effectiveness of the data governance program and data stewardship efforts</li> </ul>		<ul style="list-style-type: none"> <li>Develop the knowledge of business analysis, project delivery planning, and standards across the project life-cycle</li> <li>Understand the role and importance of the business policy analyst</li> <li>Acquire a solid understanding of the business policy analyst that encompasses business policy analysis</li> <li>Learn how to plan, elicit, analyze, model, validate, document, and manage user requirements through the project life cycle</li> <li>Conduct business policy analysis and research, provide strategic advice and research solutions to stakeholders within the organization</li> <li>Be able to ask the right analysis questions, understand and interpret the findings, and apply analysis of insights into business operations</li> <li>Effectively communicate findings and present results from business policy analysis to stakeholders</li> </ul>	
Skills	<b>Technical</b> <ul style="list-style-type: none"> <li>Mathematical statistics</li> <li>Programming &amp; Software Skills (Python, R, SAS, etc.)</li> <li>Advanced data analysis (machine learning, machine learning, text mining, and natural language processing, etc.)</li> <li>Data management</li> <li>Data visualization and presentation</li> </ul>			<ul style="list-style-type: none"> <li>Data analysis and reporting</li> <li>Data management</li> <li>Data integration</li> <li>Data quality assurance</li> <li>Data performance optimization</li> <li>Data modeling</li> <li>Data architecture</li> <li>Project/program management</li> </ul>		<ul style="list-style-type: none"> <li>Requirements elicitation and documentation</li> <li>Requirements analysis and modeling</li> <li>Business analysis tools and techniques</li> <li>Policy, research, and synthesis</li> <li>Quantitative and qualitative data analysis</li> </ul>	
Non-Technical	<ul style="list-style-type: none"> <li>Business domain knowledge</li> <li>Business ethics and governance understanding and dealing with a business situation to deliver better decision making</li> <li>Problem solving</li> <li>Critical thinking</li> </ul>			<ul style="list-style-type: none"> <li>Strategic thinking, problem-solving, innovation and collaboration</li> <li>Leadership</li> <li>Communication, writing, oral presentation, data visualization, storytelling, etc.</li> <li>Stakeholder management</li> </ul>		<ul style="list-style-type: none"> <li>Business and Business Analyst Function skills</li> <li>Business Analysts Planning and Monitoring</li> <li>Business Analysts Tools &amp; Techniques</li> <li>Planning, Eliciting, Developing, Validating, and Managing User Requirements</li> <li>Quantitative and Qualitative Policy Analysis</li> <li>Policy Writing &amp; Briefing</li> <li>Preparing for the BIA/CBA (Certification of Competency in Business Analysis) (BIA/CBA) (Certified Business Analyst Professional) Certificate Program</li> <li>Criticality, Critical Thinking, and Problem Solving</li> <li>Data Visualization and Communication</li> </ul>	
Curriculum	<ul style="list-style-type: none"> <li>Data Science Fundamentals</li> <li>Introduction to Programming</li> <li>IT/Computer Science Model Development</li> <li>Business and Technology</li> <li>Managing the Team - R / Python</li> <li>Data preparation</li> <li>Exploratory Data Analysis</li> <li>Advanced Data Analysis <ul style="list-style-type: none"> <li>Machine Learning</li> <li>Data Mining</li> <li>Statistical Intelligence</li> <li>Big Data</li> </ul> </li> <li>Criticality, Critical Thinking, and Problem Solving</li> <li>Data Visualization and Communication</li> <li>Data Provisioning &amp; Integration**</li> </ul>			<ul style="list-style-type: none"> <li>Data Governance Fundamentals</li> <li>Data Governance for Data Stewards</li> <li>Data Governance for Business Leaders</li> <li>Data Quality Assessment</li> <li>Data Analytics and Modeling</li> <li>Data Security</li> <li>Privacy Act and Regulation</li> <li>Information Management Fundamentals</li> <li>Data Governance Tools and Techniques</li> <li>Criticality, Critical Thinking, and Problem Solving</li> <li>Data Visualization and Communication</li> </ul>		<ul style="list-style-type: none"> <li>Business and Business Analyst Function skills</li> <li>Business Analysts Planning and Monitoring</li> <li>Business Analysts Tools &amp; Techniques</li> <li>Planning, Eliciting, Developing, Validating, and Managing User Requirements</li> <li>Quantitative and Qualitative Policy Analysis</li> <li>Policy Writing &amp; Briefing</li> <li>Preparing for the BIA/CBA (Certification of Competency in Business Analysis) (BIA/CBA) (Certified Business Analyst Professional) Certificate Program</li> <li>Criticality, Critical Thinking, and Problem Solving</li> <li>Data Visualization and Communication</li> </ul>	

## Analytics Program Oversight

There are a wide variety of analytical stakeholders and projects across the organization that has resulted in a patchwork of capabilities across the Department. The CDO will add value by playing an oversight and leadership role, with IT and Business, to manage, mature and optimize the analytics capability in the Department.

## Examples of Support Services

- Help you build an analytics team
- Help you find skilled analysts
- Help expand your analytics toolbox of techniques
- Help you connect to Department tools & data
- Help you access training material and training options for tools
- Help you develop/procure new tools & enhance existing solutions
- Help you get started on first project
- Design & complete projects
- Networking & collaboration opportunities



## Range of Need:

### Start:

I'm starting from scratch but want to build capacity to be self reliant.

### Grow:

I have base expertise but want to grow.

### One time:

I need help but not feasible to build analytics capacity.

## Community of Practice

The mature state of the Analytics Program envisions a strong, department-wide community of practice for analytics. This includes data scientists of course, but also identifies a number of other key players (i.e. analytics consumers, business analysts, data provisioners, apps developers). This environment will foster collaboration, knowledge and resources sharing, open practices, and provide a number of other benefits. The more employees that understand AI and related digital technologies the more diverse the voices, eyes and ears can be present to flag potential risks and impacts.

## 5.4 The Need for Internal AI Capacity at ESDC

Artificial Intelligence and related digital technologies are unique from traditional technical solutions because much of AI advancement has been because of collaborative development between multiple independent contributors. Open-source software has generated an increasingly more diverse scope of design perspective than any one organization is capable of developing and sustaining long term. This presents an opportunity for the department because to achieve true value in AI development it is critical that we have internal capacity that continues to learn and adapt to changes in the industry.

The need for internal AI capacity was one of the critical lessons learned during our early AI pilot projects. Internal capacity provides:

- The ability to develop custom solutions to save staff time and enable new ways of doing things.
- What we learn, we can leverage from one project to the next. New ideas open up as internal staff skill-up and become more familiar with our business processes.
- We're able to properly evaluate proposed AI work by external vendors (knowledge is power). This enables us to make the right business decision on behalf of Canadians.
- Enables us to demystify AI for the department, enabling more effective discussions and pragmatic AI solutions.

We know we can't build it all, especially while AI remains a relatively young industry as far as private sector solutions go. Also given the nature of Open-source software development, value can only be realized for Canadians through ESDC having its own internal capacity in AI. The department will proactively train and educate its workforce; recruit top talent; and engage with private enterprise, academia, citizens, as well as other government departments and jurisdictions.

### ESDC's Machine Learning Seminar

Since December 2016, the Data Science Division of the Chief Data Office has been hosting a weekly machine learning seminar at the ESDC Learning Centre.

Year 1 featured lecture-style presentations that introduced concepts like machine learning paradigms, deep learning and their many applications to the department.

Year 2 has been more projects focused, where seminar participants actively work together to solve AI problems:

- Several text classification solutions that identify when toxic comments are present in an internet forum
- Developed an automated scraping/analysis tool for the Canada Public Servants subreddit



- Actively working on an information retrieval project for the canada.ca website
- Methods and applications of reinforcement learning

In addition to ESDC staff, we've had attendance from several other organizations as well, including Statistics Canada and the Department of Justice. The seminar continues to evolve, but enjoys an enthusiastic, hard-working environment that is driven by its energetic membership. All ESDC staff are welcome!

Please check out our [GitLab page](#) for past topics, code repositories and more.

## 5.5 Obtaining Maximum Value with Vendors

As with any area of public service life, getting maximum value for the taxpayer should be a primary goal. As we are at the dawn of a new era of investment with respect to AI, there is an invaluable opportunity right now to properly set precedents with AI vendors on behalf of the Canadian public. The open source and collaborative nature of AI advancements presents special considerations when procuring these technologies.

Most of government procurement has operated in an environment of static technical requirements and the purchasing of solutions to well-known problems. However, AI and related digital technologies present a different reality, where development often occurs in a modular technical design process. Problems may be known but the project is completed through multiple development cycles known as iterations. Each iteration is reviewed and critiqued by the team where insights are gained and used to determine what the next step should be in the project. The solution then becomes the sum of the iterative cycles.

This reality has critical implications for how the department should consider procurement decisions that impact elements of contract management such as intellectual property, statements of work, goods vs. service determination and more. If a solution is going through multiple iterations, some done internally and others done with a vendor, will we have control over all aspects of such a development process? Particularly, the intellectual property of the tools used or results obtained, to avoid any risk of losing valuable assets such as an algorithm that can be adopted in other projects.

With these considerations in mind, we put forward vendor-guiding principles that will strategically define the manner in which we think about, assess and view our future vendor relationships when procuring AI solutions.

### *Vendor Guiding Principles*

1. ESDC's data is valuable; it should not be considered a by-product of service delivery or program policy. Its value (and access to it) should be managed with vendors accordingly;
2. If we pay the development costs for a solution, we should not be paying perpetual access fees;

3. If we understand how the technology works and how much it would cost to build it in house, we are in a much better contracting position;
4. If we have full access to the underlying algorithms and models, we can leverage them for other purposes;
5. If we're flexible in the design, we can move components around and won't be forced to continue with arrangements that no longer work for us;
6. There are hundreds of AI vendors. There is only one ESDC.

The key of these guiding principles is to provide future flexibility to the public in how it might choose to democratize the benefits of AI. That choice cannot be made for purchases that are closed one-offs, that cannot be leveraged for other projects and for which the department is paying perpetual access fees.

### ***Knowledge is Power***

All other initiatives outlaid in this strategy drive towards putting ESDC in a better position for its AI purchases:

- An effective communication strategy results in an extremely well-informed organization that knows exactly the worth of what it is buying.
- A strong internal capacity is needed to undertake the work ourselves if the right deal isn't out there, and to verify that vendors are delivering what they promised.
- Proper infrastructure and processes are needed to provide flexibility in solution deployment.

### ***Government of Canada Procurement Processes***

As AI, governance matures so will many of our internal processes and standards related to procurement, contracting and IP reflecting the policy considerations discussed in section four. The current procurement structure at ESDC does not effectively support an iterative design process. This issue is not unique to our department, but is something all GoC departments are confronted with.

To address this issue Public Services and Procurement Canada (PSPC), together with the Treasury Board of Canada Secretariat (TBS), established a list of suppliers who can provide the Government of Canada with responsible and effective AI services, solutions and products.

### ***Considerations and Risks***

It is a positive development that the government is providing more and more innovative procurement options for departments to take advantage of when developing AI solutions. However, ESDC must approach these pre-qualified arrangements with caution because the contracts may not fit the needs, requirements and legal obligations of the department.

When deciding to use streamlined procurement vehicles, we must keep in mind the above ESDC vendor guiding principles and be aware that when entering these arrangements:

- The business area should still be going through the AI governance process
- ESDC may not be the contracting authority and should not be dealing with the vendor directly in the case of contract issues
- We must consult relevant stakeholders such as the IP Center of Excellence and legal services branch if changes need to be made to the pre-qualified statements of work
- We are still governed by OSS are licensing conditions discussed in 4.2i.

## 6. Links to Other Departmental Initiatives

### Data Strategy and Analytics Program

As the current wave of artificial intelligence is largely data driven, and the techniques are grounded in Data Science, ESDC has also begun to orient itself towards AI through other initiatives. The (link to be added)ESDC Data Strategy, being led out of the Chief Data Office, aims to maximize the impact of our enterprise data asset, through modernizing how the Department approaches data governance, data management, data access, security and privacy.

Further, the (link to be added)Analytics Program, a key component of the Data Strategy, will form and scale up a frame for a burgeoning ecosystem of analytics across ESDC, to exponentially increase the insights we derive from our data. The Analytics Program includes initiatives for enhancing departmental capacity in delivering analytics solutions, enabling our staff with modern infrastructure and technology for analytics endeavors, and putting in place proper processes and oversight for deriving robust, high quality insights.

### Service Transformation Plan

The Department developed the Service Transformation Plan (STP) to support its move from strategy to implementation for transformation and modernization of its services as it moves forward in advancing it's vision for improved service delivery to clients. The Service Strategy is the departmental modernization plan of action that will transform the way we deliver service so that, in the future, Canadians will be able to digitally self-serve, access services seamlessly, receive high-quality, timely, accurate services, have their needs anticipated and receive service from a well-equipped, knowledgeable workforce.

The solutions in the Service Transformation Plan are organized in five groupings based on their impact on clients and similar capabilities:

**Allow Me** Allow citizens and clients to access their services/benefits in a faster and efficient manner.

**Trust Me** Enable better ability for clients to apply for benefits/services faster by leveraging know data about the client. Clients will feel trusted and recognized.

**Tell me** Give more information about the benefits and services and have multiple means of efficiently communicating.

**Hear Me, Show Me** Increased ability for clients to provide feedback and answer their questions.

**My Choice** Provide multiple options to engage with ESDC so clients have their choice in how they want to interact and receive benefits/ services.

More information on the Service Transformation Plan can be found [here](#).

(Placeholder for a blurb introducing STP, and see what links are available that we can direct readers to for further reading)

Several solutions in STP are actively investigating the use of AI to enable superior service. Some examples include:

- Solution 2.4: Document Upload, which will provide flexibility to clients in how they provide us with information, and is exploring the use of Computer Vision AI to automatically get the needed information into our systems.
- Solution 3.5: Chatbot, which will provide clients with the ability to first interact with a digital agent to resolve issues, speeding up resolution times.
- Solution 4.2: Program Knowledge Repository, which is investigating the use of modern information retrieval techniques (smart search) to retrieve information from our websites, manuals and other information stores in an automated, efficient manner.

### **Integrated Service Management**

At a minimum, can link Russell Egan's blog post about leveraging the power of digital and using assistive technologies such as voice-over technology, text over images, video captioning, all of which are powered by AI.

### **Putting it all together**

The initiatives outlaid in this strategy will provide critical support to all of these initiatives that intend to leverage AI, to ensure the department and its clients can realize broad-ranging, long-term value from these investments.





# **Project Charter**

## **Audit – Phase II**

### **Version 1.0**

**Date: November 5, 2018**

# 1. Charter Introduction

## 1.1 Document Change Control

Revision Number	Date of Issue	Author(s)	Brief Description of Change
1.0	2018-03-16	Nicolas Vincent, Liam Peet Pare	Document creation

## 1.2 Executive Summary

- The Chief Data Office and IASB
- The project was initiated in November 2017
- CDO was asked to come in and help in the negotiation and dealings with the vendor
- In May 2017, the CDO took over the project from the vendor to improve and finalize some of the functionalities
- Important aspects of the project:
  - Create an Artificial Intelligence enabled tool to support the classification of risk in audit reports and enable auditors to interact with them in a more efficient way. .
- Key Deliverables
  - An efficient tool to classify risk in audit reports
- Key Risks
  - Model is not able to produce results as goods as currently produced

Public Safety refuses to deploy model

# 2. Project Overview

## 2.1 Project Summary

ESDC manages a high volume of risk information spanning from branches/regions, functional groups and programs/services. It is time-consuming and resource intensive to analyze and research risks from various sources within the department.

A centralized perspective of risk intelligence is fundamental to risk management activities within the department.

The Audit team has contracted a private company to build a tool which would use machine learning techniques to extract key insights in audit reports.

The Chief Data Office (CDO) has provided some resources to assist the audit team in the negotiation of the contract and assistance in overseeing the development of the project.

First, the CDO has been successfully able to negotiate the intellectual property of the models developed. Since the department had the internal capacity to build the solution, the source code has been requested as a deliverable from the contractor. The source code has been leveraged internally to build upon and knowledge has been acquired in the methodology which has been applied in other projects.

Secondly, the CDO has been a great ally to the Audit team in the oversight of the project since it was able to make links between the technical aspects of the project and the business needs of the department. The audit team consulted the CDO throughout the whole project and together were able clear at multiple occasions the confusions that would come up during the project.

## 2.2 Project Goals, Business Outcomes and Objectives

Leverage the CDO's expertise in artificial intelligence to build an information retrieval tool for IASB. Through this project, IASB will increase its capacity to identify risk in reports, increase its analysis capacity and will be more consistent.

This project will also affect key external stakeholders that have a similar function in other GoC departments. The solution developed for ESDC could be deployed in other departments as interest was already shown in acquiring this model.

No.	Goals	Objectives	Business Outcomes
1	Enhance the search tool	Add CASA look and feel	More functionalities
2	Enhance the search tool	Provide search functions for the full documents	IASB can increase its analytics capacity while and provide consistent results
3	Report retrieval	Deploy an operational tool in ESDC	Faster and more consistent

## 2.3 Project Scope

*In scope:*



## Project Deliverable 1: CASA Look and feel and filter search

**Description:** Enhance the search functions to allow for full document search.

**Acceptance Criteria:**

**Due Date:** TBD

## 2.6 Project Cost Estimate and Source of Funding

### 2.6.1 Project Cost Estimate

### 2.7 Dependencies

- This is the second phase of this project. In phase three, the model will get significant improvements allowing it to be used more broadly across the GoC
- To develop further functionalities, more resources will need to be allocated to this project.

**Dependency Description Critical Date Contact**

## 2.8 Project Risks, Assumptions, and Constraints

### 2.8.1 Risks

No.	Risk Description	Probability (H/M/L)	Impact (H/M/L)	Planned Mitigation
1	Model fails to search through the full documents	L	M	Delayed implementation
2				
3				

### 2.8.2 Assumptions

The following table lists the items that cannot be proven or demonstrated when this Project Charter was prepared, but they are taken into account to stabilize the project approach or planning.

No.	It is assumed that:
1	IASB will allocate resources to to move this project forward
2	
3	

### 2.8.3 Constraints

*Identify the specific constraints or restrictions that limit or place conditions on the project, especially those associated with the project scope (e.g. a hard deadline, a predetermined budget, a set milestone, contract provisions, privacy or security considerations, etc.). It will help to categorize the constraints if there are several. Add rows as required.*

The following table lists the conditional factors within which the project must operate or fit.

No.	Category	Constraints
1	Hardware / software	The project requires specific hardware and software environment
2	Resources	
3		

## 3. Project Organization

### 3.1 Project Governance

Working group: IASB, CDO, **3.2 Project Team Structure**

- CDO:
  - Jeff Carr (Director)
  - Simon Harvey (Manager)
  - ~~Bjerk Ellefsen~~ / Oana Ciobanu (Senior Data Scientist / Technical Lead)
  - Wassim Athimni (Data Scientist)
  - Julia Conzon (Data Scientist)
  - Liam Garnet Peet Pare (Data Scientist)
  - ~~Nicolas Vincent (Data Scientist / Project Coordination & implementation strategy)~~
- The CDO's data scientists meet on a daily basis to discuss the progress of this project. On a regular basis the technical lead, project coordinator, manager and director are met to discuss progress.
- IASB and the CDO agreed to meet on a regular basis (weekly) to discuss the progress of the file.
- IASB:
  - Dean Shivji (Director)
  - Lorne Powell (analyst)

### 3.3 Roles and Responsibilities

- CDO
  - Liam Peet Pare - Model development



- ~~Bjerk Ellefsen~~ Oana Ciobanu - Technical Lead / Project Manager
- Nicolas Vincent - project coordination, planning / Project Manager
- Julia Conzon/~~Liam Garnet Peet Pare~~ - Model development (support)
- IASB
  - Dean Shivji – Business Lead, Objectives definition and model refinement
  - Lorne Powell – Business contact for model refinement and information acquisition

## 4. Project References

More information concerning this project can be found in the following documents:

### Document Title

Project Charter Guide

### Version #

1

### Date

November 6<sup>th</sup>, 2018

### Author and Organization

Chief Data Office

### Location (link or path)

U:\SP-PS\DMD\Data Science\02-OtherAnalyticsProjects\2018-Audit

## 5. Glossary and acronyms

### Term/Acronym

### Definition

### ***Checklist for reviewing your Project Charter:***

*After you have completed filling in the template for your Project Charter, use the list below to review the different sections to make sure you have included all the information required.*

- *The executive summary demonstrates a clear alignment between the project, the Departmental Investment plan, and the Program Activity Architecture.*
- *There are specific and measurable project objectives, as well as, business outcomes that are linked to project goals.*

- *The scope of the project is clearly stated: the reader can easily understand what product, service, or result will be delivered by the project and what high-level activities will be performed.*
- *The deliverables are spread over the duration of the project, following a phased approach composed of decision gates.*
- *Summary cost estimates and source of funding to produce internal and external deliverables are provided, including the project management and administrative effort as well as any equipment required (hardware, software, floor space, etc.).*
- *Strategic risks are identified and assessed.*
- *A governance process is defined to escalate issues when required, to approve changes to the project (scope, budget, schedule), and to accept deliverables.*
- *Authority relationships between team members are clearly presented.*
- *Project roles and responsibilities are defined and assigned to individuals or groups.*
- *Requirements for facilities and resources are described where significant logistical effort and/or funding are involved.*

*If all of these are checked as complete, then delete this checklist; update the Table of Contents and save the document to file.*